UNIVERSITAS KRISTEN INDONESIA

INTRODUCTION TO PSYCHOLINGUISTICS

COMPILED BY:

Dr. Lamhot Naibaho, S.Pd., M.Hum

UKI PRESS
Pusat Penerbit dan Pencetakan
Universitas Kristen Indonesia
Jl.Mayjen Sutoyo No. 02 Cawang
Jakarta Timur 13630

ISBN 978-623-8012-23-7

Page Cover

INTRODUCTION TO PSYCHOLINGUISTICS

Compiled By
Dr. Lamhot Naibaho, S.Pd., M.Hum

UNIVERSITAS KRISTEN INDONESIA
English Language Education Study Program
Jakarta, 2022

Publishing Information

INTRODUCTION TO PSYCHOLOGUISTICS

Compiled By
Dr. Lamhot Naibaho, S.Pd., M.Hum

Language Editor:
Matthew Kenneth Miller
Gaanesh C Waagh
Geby Arni Siregar

ISBN: 978-623-8012-23-7

Publisher: UKI Press
Member of APPTI
Member of IKAPI

Editorial: Jl. Mayjen Sutoyo No.2 Cawang Jakarta 13630
Telp. (021) 8092425

First Printed Jakarta: UKI Press, 2022

FOREWORD

A deepest gratitude and I have no other words to say other than to say thank you to GOD the Almighty, for His grace and gift, the book entitled " LINGUISTICS 1" has been compiled and published successfully.

However, in the end, I admit that this article has several shortcomings and is far from perfect, as the saying goes "there is no ivory that is not cracked" and that perfection belongs only to God. Therefore, I am happy to openly accept various criticisms and suggestions from readers, this is certainly very necessary as part of our efforts to continue to make improvements and improvements to further works in the future.

Finally, we would like to express our gratitude to all those who have supported and contributed to the entire series of processes for the preparation and publication of this book, so that this book can be presented before the readers. Hopefully this book is useful for all parties and can contribute to the development of science in Indonesia.

September, 2022

TABLE OF CONTENTS

CHAPTER 1
PROPERTIES OF SPOKEN LANGUAGE PRODUCTION

INTRODUCTION

In psycholinguistics, language production is the production of spoken or written language. It's describes all of the stages between having a concept, and translating that concept into linguistic form. According Carrol (2008:193), "language production is a fundamentally more difficult subject to study than comprehension, because although speech is observable, the ideas that lead to production are more elusive. Researchers have responded to this dilemma by using convergent measures. Some investigators have made detailed and systematic analyses of naturally occurring errors of production, and others have given speakers under laboratory conditions, more or less specific instructions on what to produce". Someone can produce the language, when they have an idea. She or he puts into words and utterances, then someone else hear the sound, recognize the words, and understand the speaker's intent. Based on the defenition above, language production can be conclude as a concept in psycholinguistics that describes the stages of speech from initial mental concept to the spoken or written linguistic result. Another word, language production is the process of communicating through language. For example; Picture a person thinking of an apple, then using their mouth to physically sound out the word 'apple', then another person hearing this and perceiving the sounds that make up the concept of 'apple', then they have the mental representation of the apple. Based on the example above, we can conclude" if the example above is an example of language production or process of communicating through language".

In language production, there are three major steps that well known as Levelt's model production that are; deciding what to express (conceptualization), determining how to express it (formulation), and expressing it (articulation) (Levelt, 1989). Researchers embracing modular models (Garrett, 1984, 2000; Laver, 1980; Levelt's, 1989, 1993; Levelt et al. 1999) have postulated the existence of a number of encapsulated, specialist modules or processes through which production process, without interaction existing among them.

A. Properties of Spoken Language Production

Although a common caricature of speaking is that it is the reverse of listening, language production processes fundamentally differ from comprehension processes in many respects. Whereas people typically recognize the words in their native language quickly and automatically, the same words require an intention to speak and can take over five times longer to generate than to recognize. For example, listeners begin to direct their gaze to the referent of a spoken noun (even in the absence of highly predictable speech) before the speaker completes articulation of the word (e.g., Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995), whereas speakers typically take about 900 ms to begin to generate a noun in isolation based on a pictured object (e.g., Snodgrass & Yuditsky, 1996).

Language production is logically divided into three major steps: deciding what to express (conceptualization), determining how to express it (formulation), and expressing it (articulation; Levelt, 1989). Although achieving goals in conversation, structuring narratives, and modulating the ebb and flow of dialogue are inherently important to understanding how people speak (for review, see Clark, 1996), psycholinguistic studies of language production have primarily focused on the formulation of single, isolated utterances. An utterance consists of one or more words, spoken together under a single into national contour or expressing a single idea (e.g., Boomer, 1978; Ferreira, 1993). While Ferreira and Englehart's chapter in this volume on syntax describes processes that allow speakers to produce their words in grammatical utterances, we focus instead on the processing of the words themselves. Indeed, most theories of multi-word utterance or sentence production ultimately boil down to an account of how sentences acquire their word orders and structures, how the dependencies between words are accommodated (e.g., subject–verb agreement), and a functionally independent account of how individual content words are generated (e.g., Chang, Dell, Bock, & Griffin, 2000; Ferreira, 2000; Kempen & Hoenkamp, 1987). In this chapter, we describe the basic properties of spoken word production, outlining empirical data that demonstrate the properties of the processes resulting in speech and discussing the processing assumptions that models of language production invoke to account for these properties. Although it could easily fill a chapter of its own, we conclude by discussing timing in multi-word utterances.

B. Generating words

Generating a word begins with specifying its semantic and pragmatic properties. That is, a speaker decides upon an intention or some content to express (e.g., a desired outcome or an observation) and encodes the situational constraints on how the content may be expressed (e.g.,polite or informal speech, monolingual or mixing languages; see Levelt, 1989). This process, termed conceptualization or message planning, is traditionally considered pre-linguistic and language neutral (Garrett, 1975; Levelt, 1989). However, speakers may include different information in their messages when preparing to speak different languages (see Slobin, 1996, on thinking for speaking). The next major stage is formulation, which in turn is divided into a word selection stage and a sound processing stage (Fromkin, 1971; Garrett, 1975). Deciding which word to use involves selecting a word in one's vocabulary based on its correspondence to semantic and pragmatic specifications. The relevant word representation is often called a lemma (Kempen & Huijbers,1983), lexical entry, lexical representation, or simply a word, and it marks the presence of a word in a speaker's vocabulary that is capable of expressing particular semantic and pragmatic content within a particular syntactic context. Sound processing, in contrast, involves constructing the phonological form of a selected word by retrieving its individual sounds and organizing them into stressed and unstressed syllables (phonological encoding) and then specifying the motor programs to realize those syllables (phonetic encoding). The final process is articulation, that is, the execution of motor programs to pronounce the sounds of a word.

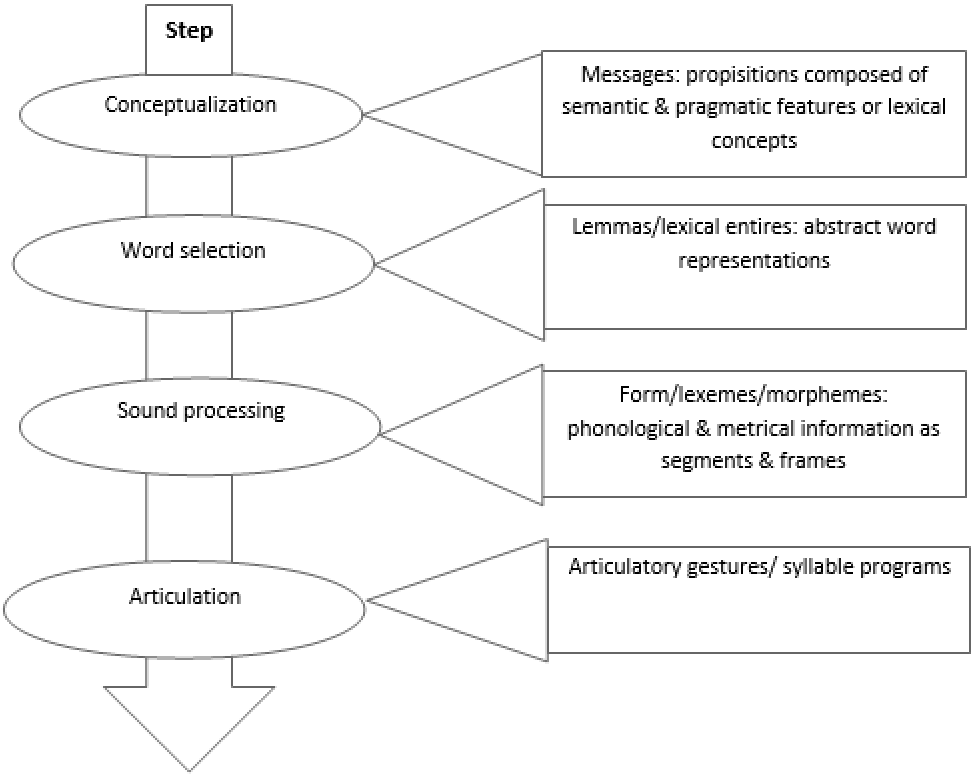


Figure 1: Major steps and representations in language production

This gross analysis of language production serves to illustrate the complexity of expressing an idea in words. The challenges posed by this complex problem give rise to the fundamental properties of the word production process, the descriptions of which below form the bulk of this chapter. Roughly, these properties delineate steps in processing (Property 1), describe how speakers deal with the relationship between meaning and word (Properties 2–9), explain how speakers represent and assemble the sounds of words (Properties 10–13), and how these processes play out in time (Properties 14 and 15). Interestingly, current models of word production agree on the basic facts about how the system works to a surprising extent, with only minor variations in explanatory mechanisms. When models differ, the tendency is to focus on different stages of production, such as word selection or phonological encoding, and different aspects of these stages such as speed of processing or how processing may go awry to yield speech errors. This means that of the properties of production described below, most are accounted for (at least to some level of detail) by most models of production. Other properties of production have yet to receive detailed attention, although we feel that much can be gained if these properties are addressed in future theories.

1. Basic Steps of Word Production

Property 1: Word selection precedes sound assembly

When speakers access word representations, they do so first based on meaning and then focus on assembling their sounds. Several sorts of evidence suggest this. The first and strongest evidence comes from analyses of errors made during spontaneous speech (Fromkin, 1971; Garrett, 1975), which reveal that speech errors most frequently involve units that can be most conservatively considered to correspond to whole words, morphemes (i.e., minimal units of meaning such as cran and berry in cranberry), or individual speech sounds (i.e., phonemes or segments such as the b- and oo-sounds in boo). In particular, error patterns suggest that a speaker may err in selecting a word but correctly assemble and pronounce its component sounds, or they may successfully select a word that can express an intended meaning, but then err in assembling its sounds. Table 1 lists examples of word, morpheme, and sound errors. In addition, the word production process occasionally falters at a point where speakers seem to have selected a word to express what they want to say but have not yet retrieved all of its sounds (see Property 8 for details).

Table 1. Speech error examples with error in bold, correction in italics, and [intended words] in square brackets.

Error type	Transcription	Source
Semantically related word substitution	"and Robbie delivers the blow that might give him the **yel-** er the **pink** oh dear let's start again might give him the *green* jersey"	Phill Liggett, Outdoor Life Network Commentator, 47 minutes into stage 13 of the 2005 Tour de France, July 15, 2005
Blend	" J**usti:ce**, *Justin and Travis*	Dr.Arthur D.Fisk in Cognitive Aging Brown Beg December 4, 2001

Phonologically related word substitution	" The battle for the green **journey** [jersey] currently on ice, until the next sprint perhaps"	Phill Liggett, Outdoor Life Network Commentator, 1st sprint on stage 21 of the 2005 Tour de France, July 24, 2005
Morpheme substitution	"It's much more useful on occasions like that. It gives the rider and the coach and awful lot of use**less** information. *useFUL information*"	Paul Sherwen, Outdoor Life Network Commentator, 1:45 into stage 19 of 2005 Tour dr France, July 22, 2005
Sound substitution	" The r**ai**d-uh. the *roadie* is accepting uh. guitar from a young man"	Participant in an experiment Griffin (2004)

That speakers produce words first by processing their meaning-level properties, then by processing their sound-level properties is arguably the fundamental property of word production. Models of word production incorporate this property by assuming two stages in producing words as well as separate word-level and sound-level representations (e.g., Caramazza, 1997; Dell, 1986; Levelt et al., 1999; but see Starreveld & La Heij, 1996, for an exception). Meaning-level representations make lexical-level representations available, which in turn provide access to individual sound/segment representations. This implies that it is not possible to go from meaning to sounds except through a lexical representation.

2. Selecting a Content Word

Despite the apparent simplicity of the resulting utterance, production of a single word can go awry in a number of ways and can take a surprisingly long time under some circumstances. Studies of isolated word production have focused primarily on nouns (e.g., person, place, or thing) with some studies of verbs (i.e., action words and predicates), ignoring other grammatical classes of content words that are less often spoken alone. In one-word utterances, the properties of word production processes appear similar for nouns and verbs (e.g., MacKay, Connor, Albert, & Obler, 2002; Vigliocco, Vinson, Damian,

& Levelt, 2002). There is no reason to suspect that other types of content words are prepared differently in single word production.

Property 2: The intention to produce a word activates a family of meaning-related words

Speech error analyses suggest that the most common error in word selection occurs when a speaker substitutes a semantically related word for the intended one, such as calling a van bus (Dell et al., 1997). A related type of speech error is a blend in which two words that could sensibly fill a particular slot in an utterance are spliced together to form an unintended string of sounds, such as behavior and deportment emerging as behortment (Wells, 1951/1973). Often the words that form a blend are not true synonyms out of context but are interchangeable only within the context of the utterance (Garrett, 1975).

Such observations suggest that accessing word representations by meaning or message representations is not surgical. Specifically, the intent to produce a particular word will lead to the activation of a family of words, all sharing some aspect of the intended meaning. This leads to the issue of how meaning is represented in models of word production, and on this issue, two major theoretical positions have been staked. On one side are decompositional views of semantic representation (Bierwisch & Schreuder, 1992; Katz & Fodor, 1963). Decompositional views portray the primitives of semantic representation as being entities that are smaller than the words whose production they ultimately support.

For ease of exposition, such features are sometimes described as themselves meaningful, so that the meaning of bird1 might include HAS WINGS, HAS FEATHERS, SINGS SONGS, and the like (e.g., Cree & McRae, 2003; Vigliocco, Vinson, Lewis, & Garrett, 2004). However, decompositional approaches can equally (and perhaps more realistically) assume that lexicalizable concepts consist of organized collections of arbitrary features (bearing a non-trivial resemblance to parallel-distributed processing accounts of cognition; McClelland & Rumelhart, 1986; Rogers & McClelland, 2004) rather than semantic primitives (or other tidy features). By either account, semantic similarity arises straight-forwardly from feature overlap – lexical items are similar to one another to the extent that the semantic features that promote their use are the same. In turn, the activation of a family of words, all related in meaning, also follows straightforwardly – if to produce bird, a speaker must activate the features HAS WINGS, HAS FEATHERS, and SINGS SONGS, then other words that also share those features will become activated (e.g., airplanes, because they have wings, opera singers, because they sing songs), and to an extent that increases with the number of shared features. Furthermore, the observation that many meaning-level influences on lexical production are context-dependent (e.g., semantic word substitutions and blends typically only involve words that can be interchanged within a particular context; Garrett, 1975; Levelt, 1989) follows from the natural assumption that representations of meaning too are activated in context-dependent fashion. Only words of the same level of specificity interfere with one another in the picture–word interference tasks, which also suggests that specificity is a important feature or constraint on word activation (Costa, Mahon, iSavova, & Caramazza, 2003; Vitkovitch & Tyrrell, 1999). Theories of production that posit decompositional semantic features include Osgood (1963), Fromkin (1971), Dell (1986), Stemberger (1985), Butterworth (1989), Caramazza (1997), and Chang et al. (2000). In addition, decompositional theories have played an important role in the development of connectionist models of word processing (e.g., Strain, Patterson, & Seidenberg, 1995) and language deficits (e.g., Hinton & Shallice, 1991).

On the other side are non-decompositional views of semantic representation. The philosophical case for non-decompositional views has been forwarded most prominently by Fodor (1975). With respect to word production, the WEAVER++ model (Levelt, 1989, 1992; Roelofs, 1993) and other models (e.g., Bloem & La Heij, 2003; Garrett, 1982; Starreveld & La Heij, 1996) have adopted non-decompositional representations. According to non-decompositional views, the representational bases of words and their meanings bear a one-to-one relationship, so that the word bird is fed by an atomic meaning representation of BIRD, the word airplane is fed by an atomic meaning representation for AIRPLANE, and so forth. These atomic meaning representations are often called lexical concepts. Within such accounts, the activation of a family of words, all similar in meaning, is not quite as straightforward as it is with decompositional accounts. Specific claims as to how multiple meanings become activated have been presented by Roelofs (1992) and Levelt et al. (1999). The idea is that activating the concept BIRD activates the concept FISH because BIRD will be connected within a semantic network to the concept ANIMAL (through what is sometimes called an “is-a” link), which will then spread activation to FISH. The concept FISH can then activate the word fish. Other links through other mediating properties will cause other meaning-similar words to become active. Figure 2 illustrates these two different forms of semantic representations.

Property 3: Words that express similar meanings compete for selection

Both error patterns and studies of object naming indicate that when a speaker strives to generate the most appropriate word for a particular occasion (often referred to as target or intended words), other words with very similar meanings in the given context become available and compete for selection as well. Intruding words in semantically related word substitutions share grammatical class, taxonomic category, and level of specificity with intended words (e.g., Fromkin, 1971; Nooteboom, 1973). Also, as described above, speakers take longer to label objects, actions, or colors in the presence of semantically related distractors relative to unrelated distractors. Although associated words such as dog and bone are related in meaning and often co-occur in speech, they do not show any tendency to compete for selection (e.g., Cutting & Ferreira, 1999; Lupker, 1979). That is, competition is restricted to words that express similar meanings rather than simply related meanings.

According to nearly all models of word production, the availability or activation level of one word affects the speed and/or likelihood that a speaker will select another word. The simplest way of modeling this is to have the probability of selecting a word (or other unit) directly related to its level of activation, so that if an unintended word has too high a level of activation, it will be erroneously selected. Due to patterns of connectivity between semantically (and phonologically) related words, these are the most likely to be highly activated and selected in error. Several models use activation levels and connectivity alone to account for patterns of speech errors (e.g., Dell, 1986; Dell et al., 1997; MacKay, 1982). Although this type of selection mechanism suffices to account for error patterns, additional assumptions are needed to account for latencies.

A number of models explain differences in naming latencies with lateral inhibition between activated word representations (e.g., Cutting & Ferreira, 1999; Harley, 1993; Stemberger, 1985). The more activated one word (or unit) is, the more it inhibits (decreases) the activation of other words. Words must reach some threshold before they are selected and the time it takes for one word to win out by suppressing others is reflected in naming latencies (for discussion of selection mechanisms in localist networks, see Schade & Berg, 1992). By contrast, in WEAVER++ (Levelt et al., 1999), the timing of word selection is influenced by two factors. The first is the activation of the to-be-selected word relative to all other activated lexical representations in a response set (em-bodying what is often termed a Luce ratio; Luce, 1959). As in many other models, the more activated a word is relative to other words, the more likely it is to be selected. The other factor is termed a critical difference, whereby a lexical representation can only be selected if its activation exceeds the activation of all other representations by some minimum amount. This critical difference is important, as it implies that one alternative representation with high activation might be a more formidable competitor than two alternative representations each with half of its activation (because the one with higher activation is more likely to exceed the critical difference threshold).

Word-production models also must account for why semantically related distractor words interfere more with word production than unrelated words do. Almost every model of word production that has aimed to explain semantic interference attributes it generally to the fact that as a speaker tries to select the most appropriate word for a stimulus, the word representations of semantically related distractors are activated more strongly and so more formidably compete for selection of the alternative form than semantically unrelated distractors (for one exception, see Costa et al., 2003). Generally, this is assumed to occur because the lexical representations of semantically related distractors receive activation from two sources: the distractor words themselves and through their semantic relation to intended words. The representations of unrelated distractor words, in contrast, do not receive this latter source of activation. For example, when naming a picture of a lion, the lexical representation of tiger would be activated not only by the distractor word tiger but also by the semantic representation of LION, whereas the lexical representation of table would only be activated by the distractor word table.

All else being equal, it takes more time for speakers to generate the names of objects that have multiple context-appropriate names (such as TV/television or weights/barbells) than those that have a single dominant name (such as tooth or bomb; Lachman, 1973; Lachman, Shaffer, &

Hennrikus, 1974). Within the language of spreading activation, less codable meanings initially activate the representation of the word that is eventually produced (e.g., TV) less than highly codable meanings do and simultaneously provide more activation to at least one other word (television), which will act as a competitor. Thus, models can explain these effects of codability on word production although the effect is not usually addressed.

Property 4: Competition for selection is constrained by grammatical class and contextual features

Not just any semantically related word competes for selection with the most appropriate word to express a meaning. Critically, the competition is limited to words of the same grammatical class. That is, only nouns substitute for nouns, verbs for verbs, and so on (Fromkin, 1971; Garrett, 1975; Nooteboom, 1973). Likewise, distractor words only appear to interfere with word production when they share grammatical class, verb transitivity (e.g., Schriefers, Teurel, & Meinhausen, 1998), and other context-relevant syntactic features with the most appropriate or target word (Schriefers, 1993; Tabossi, Collina, & Sanz, 2002). Even when substituting words are only phonologically related to the word they replace, they show a strong tendency to share the same grammatical category (e.g., Fay & Cutler, 1977). Thus, constraints on maintaining the grammatical class of an intended word appear stronger than the constraints on expressing the intended meaning.

Models of word production typically invoke different processing mechanisms to impose syntactic constraints on word selection as opposed to semantic constraints. Several models posit syntactic frames in which content words are inserted after selection. The selection mechanism is blind to word representations that do not fit the slot it is trying to fill, such that only a noun can fill a noun slot (e.g., Dell, 1986; MacKay, 1982; Stemberger, 1985).

Property 5: The speed and accuracy of selection is affected by specific meaning-level properties

A number of factors related to semantic representations and their mapping to word representations affect the speed and accuracy with which a word is selected and produced. The concreteness or imageability2 of a word is one such factor (Morrison, Chappell, &Ellis, 1997). Presumably, words with concrete, imageable meanings such as vampire, wind, and pine have richer semantic representations that guide word selection more efficiently than words with more abstract meanings such as fear, sense, and spirit (De Groot, 1992). Higher imageability or greater concreteness facilitates word translation (De Groot, 1989), generating associations (Cattell, 1889), and word recall (e.g., Martin, Lesch, & Bartha, 1999). However, highly imageable words appear more prone to substitution by semantically related words, perhaps due to a tendency to share more meaning features with other words (Martin, Saffran, & Dell, 1996). Similarly, names of objects from structurally similar categories and, in particular, living things seem more error-prone than artifacts (for comprehensive review and a theory, see Humphreys & Forde, 2001).

In addition, sentence context clearly influences the speed of word selection, probably through the influence of a combination of pragmatic, semantic, and syntactic constraints. The more predictable a word is in an utterance (based on other people's attempts to guess it from context), the less likely a speaker is to silently pause or say um before saying it in spontaneous speech (Goldman-Eisler, 1958b), in laboratory settings (Goldman-Eisler, 1958a), and the faster a speaker will label a corresponding object (Griffin & Bock, 1998).

Explaining such effects on word selection within the context of decompositional views is fairly straightforward. The properties that make particular word meanings imageable or concrete also bestow those meanings with additional features and thus richer meanings (e.g., Gordon & Dell, 2002, 2003; Hinton & Shallice, 1991). Likewise, sentence contexts may increase or sharpen the features specifying the meaning to be lexicalized. In turn, these additional features should increase the activation levels of consistent word representations while activating fewer potential competitors. With respect to imageability and concreteness, non-decompositional views might take an analogous stance, but rather than propose that imageable and concrete word meanings have richer meanings, they might propose that those meanings participate in more richly interconnected meaning networks, which might selectively promote the activation of imageable or concrete word-meanings.

Property 6: Attended objects do not necessarily lead to lexical activation

In the semantic interference effect, hearing or seeing a semantically related word interferes with generating another word. This suggests that word representations might be easily activated based on any strongly associated stimulus in the environment, even objects. For example, seeing a banana might activate the word banana to some extent, even in the absence of any intention to talk about it. Indeed, several models make this prediction (e.g., Humphreys & Forde, 2001; Roelofs, 1997). In contrast to the semantic interference effect that these theories would predict, viewing a semantically related object facilitates production relative to viewing an unrelated object (e.g., viewing a banana while trying to generate the word apple; Bloem & La Heij, 2003; Damian & Bowers, 2003). This suggests that word representations may only be activated by meaning or visual representations in the presence of an intention to communicate about them. Speakers even tend to gaze directly at objects while generating words to inaccurately label them (i.e., lie about them; Griffin & Oppenheimer, 2003). So, it seems that distractors easily influence production processing via language comprehension processes but not via object recognition processes (for some exceptions that may be due to failure to focus on what to express, see Harley, 1990).

However, the data regarding name activation for ignored visual information are not completely clear-cut. At the same time as there is evidence of perceived objects failing to show any semantic interference effects (Bloem & La Heij, 2003; Damian & Bowers, 2003), other researchers have found results that suggest phonological information about ignored objects becomes available (Morsella & Miozzo, 2002; Navarrete & Costa, 2005). Specifically, speakers are faster to name an object (e.g., a cat) in the presence of a distractor object with a phonologically similar name (cap) than an unrelated name (shoe). This is problematic given that all models require that phonological activation be mediated by word representations, so that under the same conditions that one sees phonological activation of names, it should be possible to detect semantic interference just as in other situations.

The possibility that speakers only linguistically activate words they intend to speak is an important characteristic for models of word production to take account of. The solution is to restrict activation from freely flowing from semantic representations to word representations, limiting the flow to meanings within a pre-verbal message (for examples, see Bloem & La Heij, 2003; Bloem, Van Den Boogaard, & La Heij, 2004). Specifically, Bloom and La Heij (2003) propose that until a semantic-level representation reaches a threshold level of activation, it is unable to influence word representations (only other semantic representations), and that an intention to speak is necessary to achieve that threshold activation. Note that this in essence introduces a kind of discreteness to the word-production process between the levels of meaning representation and linguistic representation.

Restricting activation flow is likely to have additional consequences and potentially explain other observations. For example, selective activation may resolve the seeming contradiction that on the one hand, imageable words are produced more quickly and accurately, whereas they are also relatively more prone to word substitution errors. Specifically, if imageable words have richer meanings or participate in richer semantic networks, then when accessed, they probably lead to the activation of a wider cohort of semantically related word meanings. When that widely activated cohort of word meanings does not extend to lexical activation, the most appropriate word representation should not suffer lexical-level competition, and so the greater activation of the intended word meaning should be free to be easily selected. At the same time, if we assume that speakers sometimes select the wrong word meaning for production or fail to highly activate important features for distinguishing similar objects, then because of the wider semantic cohort, speakers should be more likely to select the wrong meaning to lexicalize or activate a wider range of competing words when generating highly imageable word meanings than less imageable meanings.

Restrictions on activation flow between semantic and word representations cannot be blindly added to any model of production. For instance, WEAVER++ requires that meaning-level representations freely activate lexical representations in order to explain lexical-competition effects. For example, if the concept FISH (as activated through “is-a” links from the concept BIRD) is unable to activate the lexical representation of fish because FISH is not in the message, lexical competition between fish and bird is not possible and the lexical competition effects described above go unexplained.

Property 7: Selecting words is subject to long-term repetition effects that resemble learning

Selecting a word to name an object or express a meaning has long-lasting effects on how easy it is to retrieve that word again to express a similar meaning. One manifestation of this is in repetition priming for naming objects that lasts over months (Cave, 1997) and retrieving the same name for different exemplars of the same type of object (e.g., multiple cars) over the course of an experimental session (Bartram, 1974). This increase in availability is also reflected in perseveratory speech errors, such as calling a giraffe zebra after correctly naming a zebra (Vitkovitch & Humphreys, 1991). Also relevant are observations of increased latencies to name objects or actions when presented with other items from the same taxonomic category (the semantic homogeneity effect; e.g., Vigliocco et al., 2002). Note that these semantic interference and strengthening effects only occur when speakers must select words to label pictures, as sentence completions, or in some other way that is based on meaning. The perseveratory effects are not produced by reading words aloud or categorizing words as names for artifacts or natural objects (e.g., Vitkovitch & Humphreys, 1991).

Generally speaking, such long-term effects invite explanations in terms of learning. This highlights a notable gap in models of word production: Unlike the subfields of word reading (e.g., Seidenberg & McClelland, 1989; Plaut, McClelland, Seidenberg, & Patterson, 1996), phonological assembly (Dell, Juliano, & Govindjee, 1993), and even sentence production (Chang, 2002; Chang, Dell, & Bock, 2006), no major model of word production has emerged that accounts for learning effects during word production. Long term repetition priming effects and the semantic homogeneity effect could be explained via automatic strengthening of connections between meaning representations (most readily imagined as semantic features) and a selected word representation whenever the generated word successfully expresses the meaning. The fact that such long-term effects occur only with conceptually mediated production is consistent with an explanation that requires a mapping and selection rather than simple activation of meaning or word representations (as in categorization and reading). This implies that expressing the same meaning with the same word should subsequently be more efficient, but expressing a similar meaning with a different word should be more difficult. The landscape of models of word production would benefit if it included an implemented and fully developed model with learning principles that could explain these kinds of effects.

Property 8: Word production can halt part of the way through the process

Some forms of brain damage cause people to experience similar word retrieval problems for common words. When trying to come up with one of these elusive words, both brain-damaged and unimpaired speakers are able to provide a great deal of general world knowledge associated with the word and, moreover, information about the word's syntactic properties, such as whether it is a count or mass noun in English (Vigliocco, Martin, & Garrett, 1999), its grammatical gender in Italian, French, or other languages with fairly arbitrary grammatical categories for nouns (Badecker, Miozzo, & Zanuttini, 1995; Miozzo & Caramazza, 1997b), and identify the correct form of the auxiliary for sought-for verbs in Italian (Miozzo & Caramazza, 1997a). Although speakers cannot say the whole word they seek, they often can identify its first letter or sound, its number of syllables, and words that sound similar (Brown & McNeill, 1966). Provided with candidates for missing words, the speaker can (with exasperation) reject unintended words and (with gratitude) identify intended ones. Diary studies indicate that in normal life outside the laboratory, speakers typically come up with the TOT word minutes or days later. In the lab, cueing the speaker with sounds from the missing word increases the likelihood that it will "spontaneously" occur to them (James & Burke, 2000; Meyer & Bock, 1992) and providing a homophone before a TOT-eliciting stimulus makes a TOT less likely to occur (e.g., eliciting cherry pit makes speakers more likely to successfully name Brad Pitt; Burke, Locantore, Austin, & Chae, 2004).

Two models of production have sought to specifically explain TOT states: WEAVER++ (Levelt et al., 1999) and Node Structure Theory (MacKay, 1987; Burke et al., 1991). TOT states starkly illustrate the architectural assumptions of WEAVER++. After having selected word representation to express a meaning, the retrieval of the next required representation, the lexeme or phonological form of the word, fails. The successful selection of the lemma representation explains speakers' confidence that they know a word to express and their ability to report the word's syntactic characteristics (which are stored at the same level in the theory), whereas the failed selection of the lexeme representation explains speakers' inability to articulate the word. Node Structure Theory takes a similar stance, except without the lexeme (i.e., a complete lexical–phonological representation). Specifically, it postulates that a (non-phonological) lexical node fails to fully activate (allow selection of) phonological information (syllables, segments, etc.) due to weakened connections between representations. According to both accounts, partial access to phonological knowledge is accounted for by claiming partial activation of phonological representations, either as mediated by a lexeme node (WEAVER++) or not (Node Structure Theory; for further discussion, see Harley & Bown, 1998). Analogous explanations within either account can also explain the similar anomic states of brain-damaged individuals, in which they can report the grammatical genders of words that they cannot articulate (Badecker et al., 1995).

3. Generating function words and morphemes

Property 9: The selection of some words and morphemes is not primarily driven by meaning

The above-described properties can be relatively uncontroversially ascribed to the production of content words such as nouns and verbs. The production of function words such as articles (a, the) and morphemes such as past tense –ed, however, seems less driven by properties of meaning and highly dependent on the grammatical and phonological properties of accompanying words. For example, in English, the forms of indefinite determiners vary depending on whether they are used with count or mass nouns (e.g., some pasta vs. a noodle); in Swedish, indefinite determiners vary with the grammatical gender of the nouns they modify (e.g., ett bord [a table] vs. en stol [a chair]); and in French, possessive pronouns vary with the grammatical gender and phonological form of the accompanying words (e.g., mon chapeau [my hat] and mon arbre [my tree] for masculine nouns but ma table [my table] and mon ampoule [my light bulb] for feminine; from Janssen & Caramazza, 2003). This dependence on other words results in gender interference effects for determiners in which presenting a distractor noun that has a different grammatical gender than the intended object name delays speech onset for a noun phrase. So, for speakers of Dutch, production of a noun phrase such as het groene huis [the green house] is delayed by seeing the word tafel [table] that takes the definite determiner de relative to seeing the word been [leg] which takes the same determiner, het (Schriefers, 1993).

Other evidence also points to an important dissociation between content- and function word production. One is that the well-known observation that in agrammatic aphasias, function word production is notably impaired, despite the high frequency and phonological simplicity of function words (for cross-linguistic review, see Bates, Wulfeck, & MacWhinney, 1991). Another is that function words participate in speech errors in different ways than content words do (Dell, 1990; Garrett, 1975; Stemberger, 1985).

Two different classes of explanation have been proposed to account for the function content word difference. The original explanation was that function word production is associated with syntactic production, in particular a stage termed positional processing (Garrett, 1975). Function words are integral parts of the positional frames with which speakers bind their to-be-produced content words. The use of particular function words is conditional upon meaning-level properties, so positional selection must be sensitive to meaning-level features. Nonetheless, such retrieval is performed differently than is done with content words, which are claimed to be retrieved in a manner like that described earlier. A second related proposal is that function words are selected after the retrieval of the content words that license their use (Ferreira, 2000; Levelt, 1989). The general idea is that a speaker might retrieve, say, a noun, which in turn will trigger the retrieval of knowledge (Ferreira, 2000) or execution of a procedure (termed indirect election by Levelt, 1989) that retrieves the needed function words. Again, this amounts to a claim that a function word is not retrieved directly by meaning, but is instead mediated by the syntactic properties of content-word knowledge.

4. Assembling the Sounds of a Word

Property 10: The sounds of a word are assembled anew

A potentially counterintuitive idea is that the individual sounds of words are assembled a new each time they are spoken rather than retrieved as intact wholes. Yet, patterns of speech errors and latency data suggest that this is the case. According to one estimate, errors involving sounds occur approximately 2.6 times per 1000 sentences or 1.5 times per 10,000 words in spontaneous speech, whereas word errors occur at a rate of 4.4 per 1000 sentences or 2.5 per 10,000 words (Deese, 1984). When unimpaired speakers name isolated objects, errors involving the sounds of words are much less likely than word substitution errors (Dell et al., 1997). Sound errors include omissions, additions, and ex-changes of individual sounds. The most common type of error is the anticipation of an upcoming sound (Nooteboom, 1973) as in alsho share for also share (Fromkin, 1971).

Level and colleagues have argued that one reason that a word's sounds must be assembled anew each time is due to changes in metrical structure contingent on the accompanying words and inflections (e.g., Levelt et al., 1999). For example, whereas the /d/ is syllabified with the other sounds in hand when the word is spoken alone, it is syl- labified with the following word, it, in the utterance Hand it ["han-dit"]. The importance of metrical structure can also be seen in benefits of repeating syllable structure independent of syllable content (Sevald, Dell, & Cole, 1995). That is, speakers can repeatedly produce kem–til–fer much faster than they can produce kem–tilf–ner, because the first two syllables of the first sequence share syllable structure (CVC) whereas the first two syllables of the second sequence do not (CVC and CVCC).

Property 11: Experience strongly affects speed and accuracy of assembling words

A striking fact about slips of the tongue is the way they reflect both long- and short- term experience with language patterns. Words fall apart in ways that reflect the sequences of sounds a speaker is most familiar with. Slips of the tongue are more likely to create words that exist in a speaker's language rather than create novel sequences of sounds, a phenomenon known as lexical bias (Baars, Motley, & MacKay, 1975; Dell & Reich, 1981). Even when novel sequences are created, sounds in these new sequences only occur in syllable positions that they occupy in existing words of the language.

The sounds that slip tend to be those in the least predictable positions within the language of the speaker (Berg, 1998; Nooteboom, 1973). For example, word initial consonants (e.g., /b/ in the word bicycle: /baj.sI.kl/) are less predictable than other consonants that begin syllables in English (e.g., the /s/ and /k/ in bicycle), and in Germanic languages, slips of the tongue are more likely to separate a word initial consonant from a word than another syllable initial consonant (Shattuck-Hufnagel, 1987). This also shows up in word games and sayings where everything except the initial consonant of a word is repeated (e.g., Pig Latin and "helter skelter" or "piggely wiggly"). Languages like Spanish that do not have this difference in distribution of word-initial vs other syllable-initial consonants do not show this difference in phonological speech errors (Berg, 1991). Whole syllables often participate in speech errors in Mandarin Chinese (Chen, 2000, as cited by Chen, Chen, & Dell, 2002) but rarely in English. Likewise, in Semitic languages such as Arabic, where morphemes are discontinuous and the syllable positions of consonants change more often across words, speech error and poetic rhymes pattern very differently from languages like English where morphemes tend to be concatenated and maintain syllabic position across words (for comprehensive review, see Berg, 1998).

Even short-term experience with particular sound-ordering conventions affects the likelihoods of making different types of errors. So, when in the context of a particular experiment, sounds only occupy particular positions, speakers' speech errors come to reflect these biases even when they are not part of the language in general. For example, when /f/ only occurs at the beginning of syllables in an experiment, speech errors involving /f/ nearly always involve the beginning of syllables (Dell, Reed, Adams, & Meyer, 2000). Even individual phonological features such as place of articulation (e.g., the difference between /b/, /d/, and /g/) are sensitive to these effects of experience (Goldrick, 2004).

Speakers appear sensitive to the frequency of whole words in addition to sequences of sounds within them. Unsurprisingly, children tend to learn common words earlier than uncommon words (e.g., Huttenlocher et al., 1991). Thus, it is difficult to determine whether it is the age at which a word is typically learned (its age-of-acquisition), how often it tends to be used (its word frequency), or both that affect word production (for a discussion of attempts, see e.g., Brysbaert & Ghyselinck, in press).5 More common or earlier learned words are generated as much as 100 ms rapidly than less common words (Oldfield & Wingfield, 1964). This speed advantage for common words may be due to the benefits of experience in word selection and phonological encoding, but several results suggest that the impact of frequency and age-of-acquisition is greater in phonological encoding than in word selection (see Brysbaert & Ghyselinck, in press). For example, lower word frequency increases the likelihood of phonological word substitutions, slips of the tongue (Dell, 1990; Stemberger & MacWhinney, 1986), and TOTs (Burke et al., 1991; Harley & Bown, 1998), but only seems to affect the likelihood of semantic word substitutions in unimpaired speakers when they are under heavy time pressure to speak (Vitkovitch & Humphreys, 1991).

Unlike the above-noted effects of the frequency of words or word patterns, the effect of the frequency of syllables upon production is less clear.

Levelt and Wheeldon (1994) reported that the frequency of the final syllables of words influences production time independently of word frequency or whole-word naming time. More recent experiments showed that for disyllabic non-word production, Dutch-naming latencies (the same language assessed by Levelt and Wheeldon) were influenced by the frequencies of first syllables but not second syllables (Cholin, Levelt, & Schiller, in press). More work is necessary to sort out exactly when syllable frequency does and does not affect production times.

Property 12: Aspects of sound assembly proceed sequentially

Processing seems to start earlier in time for sounds at the starts of words than for sounds at the ends. For example, in picture–word interference tasks, distractors that share the initial sounds in object names have effects at earlier points in time than distractors that share later sounds of names (Meyer & Schriefers, 1991; e.g., an initially overlapping word like tile facilitates naming of a tiger at earlier points in time relative to a word like liar that only overlaps in final sounds). When speakers are told in advance that they will be asked to articulate one word out of a set that share initial sounds, they begin speaking earlier than if the word comes from a group that does not share initial sounds (Meyer, 1990, 1991). That is, speakers appear able to prepare the shared parts of the words in advance, leaving less material to prepare and allowing faster production than when nothing is known in advance about the form of the upcoming word. However, this fore knowledge is only helpful in Dutch when words share initial sounds and metrical structure (specifically, number of syllables and stress pattern; Roelofs & Meyer, 1998). In contrast, knowing in advance the final sounds or syllables of words provides no benefit (Meyer, 1990, 1991). Knowing just a phonological feature such as place of articulation at the start of a word also does not provide any detectable benefit, suggesting that the relevant level of preparation involved is whole phonemes or sounds (Roelofs, 1999). Somewhat similarly, generating a word with word-initial overlap (e.g., tile and tiger) slows naming of an object on a subsequent trial, whereas generating a word with word final overlap (anger and tiger) speeds naming relative to generating an unrelated word (Wheeldon, 2003). When speakers repeat word pairs multiple times, it takes more time per pair for combinations that share initial sounds relative to those that share no sounds, which in turn take more time than combinations that only share final sounds

(Sevald & Dell, 1994). Although there are intriguingly different patterns of effects for hearing a word with overlapping initial segments vs. generating one, all of these results suggest a sequential process associated with retrieving, organizing, or programming speech sounds.

The WEAVER++ model covers all bases by having both simultaneous retrieval of all segments in a word, followed by a step in which each segment is associated with a syllable position in sequential order (Levelt et al., 1999). Thus, speed of processing in the model is sensitive to the availability of all phonemes in the first part of phonological encoding and the time needed to sequentially associate them with a syllable position. In addition, there is a final stage of phonetic encoding in which the phonologically specified syllables of words are sequentially associated with stored articulatory gestures. While having two stages that show sequential processing allows the model to account for sequential effects in production, it also makes it difficult for the model to simultaneously account for the absence of certain length effects. Specifically, object naming latencies and gaze durations on objects suggest that when the many potential confounds with word length are controlled, speakers take the same amount of time to prepare a multi-syllabic word as a monosyllabic one (Bachoud-Levi, Dupoux, Cohen, & Mehler, 1998; Bonin, Chalard, Meot, & Fayol, 2002; Griffin, 2003; Sternberg, Knoll, Monsell, & Wright, 1988; for discussion, see Meyer, Roelofs, & Levelt, 2003).

Property 13: The effect of similar sounding words is highly situation-dependent

The effect of recent experience with a word that is phonologically similar to an intended word sometimes speeds (e.g., Starreveld, 2000) and sometimes slows word production (e.g., Wheeldon, 2003), indicating that such effects depend on a complex set of factors. These differing effects may be due in part to experimental paradigms differentially calling on phonological subprocesses, such as sound retrieval as opposed to sound sequencing, and similar sounding words having different effects in sound retrieval, associating sounds with metrical structure, translating these phonological plans into motor programs, and articulation (see Levelt et al., 1999; O'Seaghdha & Marin, 2000). Complicating the interpretation of phonological priming effects in production (e.g., Starreveld, 2000), similar sounding words compete with one another in word recognition (e.g., Tanenhaus et al., 1995). In addition to the position in the words where sounds are shared (Sevald & Dell, 1994), the duration and type of processing that the first or priming word undergoes appears critical in shaping effects (O'Seaghdha & Marin, 2000).

When a speaker unintentionally asks for balaclava in a Mediterranean restaurant (rather than baklava), it is tempting to conclude that similar word forms compete against one another. Instead, speakers may say words that sound similar to their intended words as near-misses, in which they fail to retrieve all of the sound information for an intended word and default to a very similar form (Burke et al., 1991). In this category of effects, one can list the tendency for slips to be real words rather than novel sequences of sounds, the tendency for intruders in phonological word substitutions to have the same number of syllables and other characteristics as intended words (Fay & Cutler, 1977; Gagnon et al., 1997), the tendency for speakers in TOT states to often come up with similar sounding words (Burke et al., 1991), the tendency of slips of the tongue to involve sounds that share many phonological features, such as /t/ and /k/ rather than /t/ and /v/ (Fromkin, 1971; Shattuck-Hufnagel & Klatt, 1979), and the tendency for sounds to exchange between similar sounding words (e.g., Dell & Reich, 1981). That is, words that sound alike do not appear to interfere and compete with one another during phonological encoding in the way semantically related words do in word selection. Indeed, although selecting a word from a semantically dense neighborhood seems to take more time than selecting one from a sparse neighborhood, the opposite seems to hold for phonologically defined neighborhoods. Words that share many sounds with other words take less time to generate than words that are more unusual (Vitevitch, 2002) and appear more likely to be successfully retrieved in terms of fewer phonologically related word substitutions (Vitevitch, 1997) and TOTs (Harley & Bown, 1998). Also supporting the idea that similar sounding words support each other rather than compete is the observation that priming with phonologically related words can resolve and prevent TOTs (James & Burke, 2000; Meyer & Bock, 1992). Likewise, presenting phonologically related distractor words during object naming speeds naming latencies relative to unrelated distractors (e.g., liar vs. ankle for a lion; Schriefers et al., 1990). Simulation studies conducted with interactive activation models suggest that feedback of activation from phonological neighbors may aid intended words in competing against their semantic neighbors (Dell & Gordon, 2003).

That said, there are situations in which having similar sounding words slows speech or increases the likelihood of errors. The most obvious case of such interference is in tongue twisters such as The sixth sick sheik's sixth sheep's sick. Experiments indicate that repeating words with similar initial sounds is significantly more difficult than repeating sequences with unrelated sounds (Sevald & Dell, 1994). In addition, speakers are slower to generate a name for an object (e.g., hoed meaning hat in Dutch) when they generated a word with overlapping initial sounds (hond) on the preceding trial rather than an unrelated word (Wheeldon, 2003). Dell and O'Seaghdha (1992) suggested that these and related phenomena reflect sequentially cued phonological competition, whereby having completed a sequence of phonemes (e.g., /i/) with one ending (/k/, in sheik), cueing of the recently used ending makes it more difficult to subsequently complete that sequence (/i/) with a different ending (/p/, in sheep). Such sequential competition is readily accounted for with the subclass of connectionist models called simple recurrent networks and control signal networks that output phonological segments one at a time for a given input (e.g., Dell et al., 1993; Vousden, Brown, & Harley, 2000). Although such models do an excellent job of producing some phenomena associated with phonological word assembly (particularly the effects of experience, similarity, and order on speech errors), it is unclear how they would be integrated with other parts of the production system to account for phenomena such as phonological influences on word selection.

5. Time course of processes in word production

Property 14: Semantic competitors activate their sounds

Despite the near consensus on the need for two stages to the production process, a famous controversy among theories of word production concerns the extent to which processing of sound and meaning overlap in time. In one manifestation of this, researchers have debated whether sound-related information is only processed after word selection is complete (e.g., Dell & O'Seaghdha, 1992; Dell & Reich, 1981; Harley, 1993; Levelt et al., 1991; Peterson & Savoy, 1998). On one side are models that characterize the flow of information during production as strictly staged – speakers first use activated meaning-level representations to perform word selection and only access sound information after the completion of the selection process. The most prominent model of this strictly discrete sort is the WEAVER++ model presented in Levelt et al. (1999), which was developed computationally in Roelofs (1992, 1997). Other theorists have also argued for the strict separation of word selection and sound processing stages (e.g., Butterworth, 1989; Caramazza, 1997). On the other side are models that assume staged processing, but allow activation to flow relatively freely among meaning, lexical, and sound representations, making multiple types of information relevant to both word selection and sound assembly (e.g., Dell, 1986; Harley, 1993). Specifically, partially activated but ultimately unselected lexical representations are permitted to influence sound assembly (via cascading activation). For example, before ultimately naming an object as couch, a speaker should activate both the word representation for couch and its synonym sofa (see Property 2, that speakers activate a family of meaning-related words) and, via cascading, the sounds of these words. Indeed, speakers are faster to read aloud words that are phonologically related to dispreferred synonyms of object names (e.g., soda for sofa when preparing to name a couch) when they are presented after beginning to prepare to name a drawing of a couch rather than a completely unrelated object (Peterson & Savoy, 1998; replicated by Jescheniak & Schriefers, 1997; see also Jescheniak & Schriefers, 1998). The WEAVER++ model makes the post-hoc assumption that word selection is delayed until after sound processing begins only in the case of synonyms (Levelt et al., 1999).

Another manifestation of this controversy has concerned whether the sounds of ultimately unselected words may influence which word is selected. In models with bidirectional flow of activation or feedback, partially activated but ultimately unselected phonological representations are allowed to send activation backwards to affect lexical (and perhaps even semantic) levels of representation. The most prominent implemented model of this interactive sort is presented in Dell (1986), but this type of interactive activation has been incorporated in many theories and models (e.g., Dell et al., 1997; Eikmeyer, Schade, Kupietz, & Laubenstein, 1999; Harley, 1993; MacKay, 1982, 1987; Stemberger, 1985). Explaining the mixed error effect is one of the primary motivations for assuming this type of interactivity. It turns out that the intruding words in semantically related word substitutions bear a greater than chance phonological similarity to the intended words that they replace (Brédart & Valentine, 1992; Dell & Reich, 1981; Harley, 1984; Martin, Weisberg, & Saffran, 1989). In interactive-activation models with feedback, when generating the word cat, activation spreads to words related in meaning to cat such as dog, mouse, and rat, and via their word representations to their sounds. The sounds that form the word cat are highly activated by their link to cat's word node and they relay a portion of that activation to other words containing the same sounds such as cap, kit, and rat. Thus, a word that is both semantically and phonologically related to the intended word such as rat receives converging activation from both semantic and phonological representations, making it more likely to be selected by mistake than a word activated by only one of these sources. In contrast, discrete two-stage models account for mixed errors uses an independently motivated error-checking mechanism (see e.g., Motley, Camden, & Baars, 1982). The basic idea is that the more a substituting word resembles an intended word, the less likely a pre-articulatory editing mechanism is to detect the error and prevent it from being uttered. Thus, under this account, mixed errors are not made disproportionately often, it is just that other errors are more likely to be detected and prevented, making the types of errors observed unrepresentative of those created in the language production system (Butterworth, 1982; Levelt, 1989). Thorough treatments of the issues of discreteness and interactivity in word production can be found in Rapp and Goldrick (2000) and Vigliocco and Hartsuiker (2002).

Property 15: The scope of message planning is greater than the scope of sound assembly

Early in the study of speech error patterns, researchers noted that there was a greater distance between words that exchange places than between sounds that exchange places. For example, Nooteboom (1973) noted that 2.1 syllables separated exchanging sound segments, whereas 4.1 intervening syllables was the average distance between exchanging units of greater size such as morphemes and words. Such observations support the distinction between word and sound representations and separate processing stages that operate on them (e.g., Fromkin, 1971; Garrett, 1975). In addition, it suggests that abstract properties of words are specified further in advance than their sounds are.

There is a tradition in psycholinguistics of searching for units in which planning is incremented. With respect to the minimum amount of planning a speaker must complete before beginning a fluent utterance, the primary units proposed have been based on prosody or syntax. In the psycholinguistic literature, a phonological or prosodic word is typically defined as a single content word along with any adjacent, unstressed function words as in [beer's a] [good] [thing] (Ferreira, 1993; Wheeldon & Lahiri, 1997). Latency data indicate that speakers prepare at least one phonological word prior to initiating an utterance, with more complex phonological words delaying speech onset (Wheeldon & Lahiri, 1997). Other studies suggest that speakers will prepare more than one phonological word prior to speech if it does not form a whole lexical word (i.e., it is half of a compound; Wheeldon & Lahiri, 2002); if the second phonological word is part of the first noun phrase (Costa & Caramazza, 2002); and if the first word will not take long to articulate and speakers try to avoid pausing (Griffin, 2003). Strengthening the case for considering the phonological word an important unit at some level is the observation that the latency to begin articulating pre-planned speech is a function of the number of phonological words the pre-planned utterance contains (Sternberg et al., 1988; Wheeldon & Lahiri, 1997).

Other researchers have argued for phrase-wise word planning (e.g., Martin, Miller, & Vu, 2004). Certainly in many languages (e.g., Dutch, German, and Spanish), grammatical dependencies between nouns (e.g., beer, ale) and the adjectives (e.g., hoppy, amber) and determiners (e.g., a, some) that modify them make it necessary to retrieve information about the noun to determine the correct form of the adjective or determiner. Not surprisingly, picture–word interference studies suggest that nouns are selected before the onset of the determiner when speakers produce gender-marked determiner + adjective + noun phrases in languages such as Dutch and German (Schriefers, 1992; Schriefers, de Ruiter, & Steigerwald, 1999). Other work points to phrase-wise planning even in English speakers in the absence of strong grammatical dependencies. For example, English speaking patients who, because of brain damage, have difficulty maintaining lexical– semantic information had greater difficulty producing utterances in which adjectives appeared in the same phrase as the noun they modified (e.g., the long, brown hair) than utterances in which the adjectives appeared in a different phrase (e.g., the hair was long and brown; Martin & Freedman, 2001). Tellingly, patients with impaired memory for phonological information did not show this difference and could produce these utterances as readily as unimpaired speakers.

There is mixed evidence for pre-speech planning of multiple nouns when they occur in a conjoined noun phrase such as monkey and chair. Support for phrasal planning comes from finding of semantic interference effects on speaking latencies for both objects within a conjoined noun phrase (Meyer, 1996; but see Meyer, 1997) and when the nouns in the conjoined phrase name semantically related objects (Freedman, Martin, & Biegler, 2004). All else being equal, timing experiments indicate that speakers take about 70 ms longer to initiate sentences with two nouns in a conjoined subject noun phrase than sentences with a single noun (Martin et al., 2004; Smith & Wheeldon, 1999). Such observations have been used to argue that the contents of a noun phrase are processed in parallel (with a small cost) and that articulation of a sentence-initial conjoined noun phrase is not initiated until both nouns are prepared. In contrast, eye-tracking experiments suggest that under similar circumstances, speakers prepare nouns one at a time, shortly before uttering them, even in complex subject noun phrases or conjunctions (Griffin, 2001; Meyer, Sleiderink, & Levelt, 1998; but see Morgan & Meyer, 2005).

An unanswered question is whether verbs (or other predicates) play a special role in the preparation of utterances. Based on the constraining properties of verbs, some theorists have suggested that verb selection must normally take place early in sentence formulation (e.g., Bock, 1987; Ferreira, 2000; Jarvella, 1977; MacWhinney, 1987). When not required to select verbs in an utterance, speakers begin speaking earlier than they otherwise do (Kempen & Huijbers, 1983; Lindsley, 1975). Such results have been used to argue that verb selection precedes subject selection and therefore often speech onset (e.g., Bock & Levelt, 1994; Ferreira, 2000). However, these same experiments (Kempen & Huijbers, 1983; Lindsley, 1975) are also consistent with a desire to have a full or partially specified message planned before speech onset without verb selection, assuming that messages that include an action or other predicate take more time to compose, all else being equal, than those with only a topic. Similarly, the relationship between ear-to-mouth lag and verb position in translation input (Goldman-Eisler, 1972) supports the idea that a verb is selected before translated production begins, but also the more conservative possibility that production processes wait for a predicate to be included in the message. Further complicating matters is the possibility that speakers may only need to prepare verbs prior to speech onset whenever verbs occur soon after sentence onset (e.g., after short subject noun phrases in English) simply because there would not be time to prepare them while articulating the subject noun phrase (Griffin, 2003).

In addition to semantic and linguistic units and dependencies, time also appears to be important in timing speech. Longer words by definition take more time to articulate than shorter words do and slower speakers take more time to articulate their words than faster speakers do. Both of these aspects of timing have been shown to influence when speakers begin preparing words (Griffin, 2003). That is, speakers may attempt to minimize their buffering of prepared words by estimating how long words will take to prepare and how long it will take to articulate already prepared speech. Speakers are sensitive enough to the timing of articulation and word preparation that they will insert optional words such as that is The coach knew that you missed practice is response to variations in the availability of the following word (Ferreira & Dell, 2000). Also suggesting sensitivity to the time needed to prepare upcoming speech, speakers are more likely to say uh than um before shorter delays in speaking (Clark & Fox Tree, 2002).

CONCLUSION

This chapter has described 15 basic properties of spoken language production. These properties characterize word production as consisting of a word-selection stage followed by a sound-processing stage (#1). Selecting a content word such as a noun or verb involves activating (#2) and then competitively selecting (#3) from a family of meaning- related words in a grammatically constrained (#4) but meaning-sensitive (#5) fashion. This word-selection process may require an intention-to-name to have it commence (#6), and it manifests a long-term learning component (#7). Nonetheless, word production can fail partway through (#8). Function words may undergo a somewhat different selection process than content words do (#9). Sound processing in turn is characterized as assembling sequences of sounds (#10), a process that is affected by speakers' experience (#11), and proceeds from word start to end (#12). Phonological similarity has complex effects on production, attesting to the fact that it probably affects multiple subprocesses (#13). Although only one word may ultimately be spoken to produce a meaning, multiple meaning-related candidates can affect the availability of sound information (#14). Finally, speakers plan messages further in advance than they retrieve sounds, showing a tendency to prepare words for about a noun phrase at a time, due to message-level, syntactic, prosodic, and/or timing constraints or preferences (#15).

In focusing on spoken language and the production of words in particular, we have left untouched the literature on written language production (see e.g., Bonin et al., 2002; Kellogg, 2003), the production of sign languages (e.g., Thompson, Emmorey, & Gollan, 2005), and the complications of knowing words in multiple languages (e.g., Costa, Miozzo, & Caramazza, 1999; Gollan & Acenas, 2004; Kroll & Sunderman, 2003). Within spoken word production, this chapter has not addressed work on how speakers produce morphologically complex words (e.g., Badecker, 2001; Melinger, 2003; Roelofs, 1996; Wheeldon & Lahiri, 2002; for discussion, see Waksler, 2000) such as morphology, litterbox, or ko-tätaste (a Swedish word meaning "most tightly packed with cows") or idioms such as It's Greek to me and to put one's foot in one's mouth (see e.g., Cutting & Bock, 1997; Levelt et al., 1999). We have hardly touched on the production of prosody and the role of intonation in spoken language (for discussion, see Ferreira, 1993; Wheeldon, 2000). Nor have we discussed under what circumstances and how speakers may or may not tailor their language to suit their audiences (Barr & Keysar, this volume; Ferreira & Dell, 2000; Ferreira, Slevc, & Rogers, 2005; Horton & Gerrig, 2005; Kraljic & Brennan, 2005; Lockridge & Brennan, 2002). These are active and important areas of research in language production.

Most of the properties we have reviewed are sufficiently basic that they are virtually certain to characterize how production works, at least to some level of approximation. A few of them, however, are more controversial and are likely to be explicated and revised by future research (e.g., whether the intention to name is critically involved in word activation [6], seriality in phonological encoding [12], and origins of phonological similarity effects in production [13]). Nonetheless, in all, these properties represent a tribute to the progress that the field of language production has made, as they represent true gains in our understanding of how speakers produce words. At the same time, they pose challenges to current and future models of production, as such models pursue their goal of transforming these descriptions of how production works into explanations of why it works the way it does.

CHAPTER 2
SYNTAX AND PRODUCTION

INTRODUCTION

Syntax has to do with how words are put together to build phrases, with how phrases are put together to build clauses or bigger phrases, and with how clauses are put together to build sentences. In small and familiar situations, humans could communicate using single words and many gestures, particularly when dealing with other members of the same social grouping (nuclear family, extended family, clan and so on). But complex messages for complex situations or complex ideas require more than just single words; every human language has devices with which its speakers can construct phrases and clauses. We habitually talk of human languages and their speakers; we ask questions such as ‘How many speakers are there of Chinese/Arabic/ Spanish?’ Nobody ever asks how many writers such-and-such a language has, but the distinction between speaking and writing is crucial and affects the study of syntax. It is therefore surprising that we cannot draw a major distinction between spoken and written language. Instead, the major distinction is between language for which very little planning time is available and language for which much more planning time is available. Much spoken language is indeed produced with little planning time, but some kinds are planned or semi-planned. A current-affairs report on radio is written but spoken aloud, while lectures in universities have at least an outline script in the form of ‘headlines’ projected onto a screen but require some improvisation. Many types of writing involve planning, such as essays, research papers and books, but other types of xii 01 pages i-xvi prelims 18/10/01 4:49 pm Page xii written text are typically produced quickly, such as personal letters and e-mail messages to friends or close colleagues.

Many kinds of spoken language, not just the spontaneous speech of domestic conversation or discussions in pubs, have a syntax that is very different from the syntax of formal writing. It is essential to understand that the differences exist not because spoken language is a degradation of written language but because any written language, whether English or Chinese, results from centuries of development and elaboration by a small number of users – clerics, administrators, lawyers and literary people. The process involves the development of complex syntactic constructions and complex vocabulary. In spite of the huge prestige enjoyed by written language in any literate society, spoken language is primary in several major respects. There are, or were until recently, societies with a spoken language but no written language, but no societies with only a written language; children usually learn to speak long before they learn to read and write; and the vast majority of human beings use speech far more often than writing. The syntax of spontaneous spoken language has been 'designed' or 'developed' to suit the conditions of speech – little planning time, the possibility of transmitting information by loudness, pitch and general voice quality, and support from hand gestures, facial expressions and so on (what is known as 'non-verbal communication'). For a particular language, the syntax of spontaneous speech overlaps with the syntax of formal writing; there is a common core of constructions. For instance, The instructions are useless could be spoken or written. However, many constructions occur in speech but not in writing, and vice versa. She doesn't say much – knows a lot though is typical of speech, but typical of writing is Although she does not say much, she knows a lot. The special syntax of spontaneous spoken language is not produced just by speakers with the minimum of formal education. One of the most detailed investigations of spoken syntax was carried out in Russia in the late 1960s and early 1970s. The speakers recorded on tape in all sorts of informal situations were doctors, lawyers and academics, but their speech turned out to be very different in syntax from written Russian. Moreover, their syntax had general properties which have turned up in bodies of spontaneous spoken English, French and German.

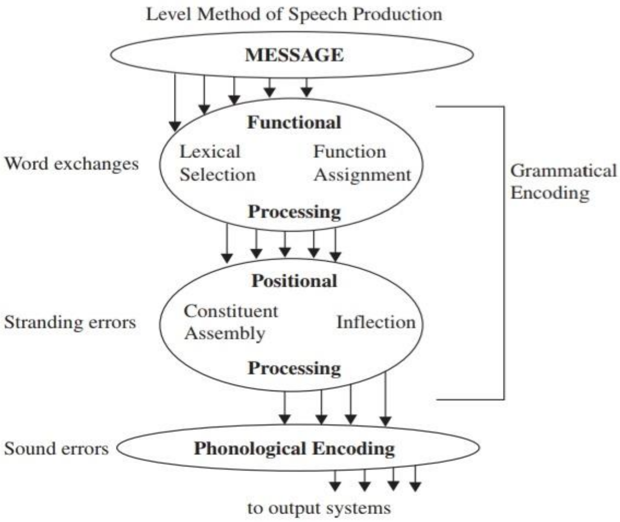
People learn the syntax and vocabulary of formal writing from books and in school in a process that lasts into the early twenties for university graduates and can continue much longer. In general, the more exposure speakers have to formal schooling, the more easily and frequently they use in speech the syntax and vocabulary that are typical of formal writing. Individuals have choices, however; a highly educated individual may choose to keep to simple language in speech and writing, and individuals with a minimum of formal education but a large exposure to books may use very complex language in all situations. The concept of a language is not straightforward. People think of themselves as, say, speakers of French or speakers of English, but they can be thought of as possessing a core of grammar and vocabulary and a greater or lesser number of other genres, possibly with special syntactic constructions but certainly with special vocabulary and fixed combinations of words; the language of literary criticism is different from the language of football reports. The concept of a language is not straightforward. People think of themselves as, say, speakers of French or speakers of English, but they can be thought of as possessing a core of grammar and vocabulary and a greater or lesser number of other genres, possibly with special syntactic constructions but certainly with special vocabulary and fixed combinations of words; the language of literary criticism is different from the language of football reports. Syntax is neutral with respect to 'correct' and 'incorrect' English, French and so on. Analysts of English aim to cover as much data as possible. They collect samples of current speech and writing and note that examples such as (1) are typical of speech but also occur in writing while examples such as (2) occur mainly in formal writing. That is, they analyse and describe all the data they come across. Language is at the centre of human societies; it plays a crucial part in the organisation of social activities, from government through the workplace to the home. These complex tasks require complex language, and that requires syntax.

A. Syntactic Representations in Production

1. Two-Stage Models of Grammatical Encoding

As has been argued since the earliest days of generative grammar, syntax is an interface between meaning and sound (articulation/phonetic form). A word such as cat has a particular meaning, but the expressive power of language is enhanced immeasurably by our ability to create meanings compositionally, by putting words together – for example, our ability to say not a cat or that's my cat. Models of production instantiate this basic architecture fairly, transparently. Consider the Bock and Levelt (1994) model, which was described in the previous Handbook of Psycholinguistics.

The model, henceforth referred to as Bock–Levelt (BL), is shown below (reprinted from the 1994 chapter):



The process of speaking begins with a message-level representation, which captures the idea the speaker wishes to convey. This message becomes sound at the other end of the model, at a stage called phonological encoding. Linking the message and phonological levels are two stages of syntactic processing (or grammatical encoding, as it is called in the model), one called functional processing, and the other positional processing. Notice that the basic linguistic architecture in which syntax mediates between meaning and form is replicated in the BL model of production. Yet an important difference is that syntactic operations are factored into two components. This two-stage architecture originated with Garrett (1975), who argued from speech error data that the production system first creates the global, syntactically functional structure for a sentence, and in a separate stage determines phrasal details such as serial order. In BL, grammatical encoding begins with functional-level processing. Abstract lexical entries termed lemmas, which contain information about a word's meaning and its syntactic requirements but do not represent its phonology, are retrieved and assigned grammatical functions such as subject and object.2 For example, for the utterance my cat terrifies the dog next door, the lemmas for CAT, DOG, NEXT, DOOR, and TERRIFY would be retrieved, and CAT would be assigned the role of subject and DOG (modified by NEXT DOOR) the role of object. At this point, then, the speaker has committed to some type of active structure in which CAT will be the subject;

a passive structure is ruled out, because in any type of passive, DOG would be the subject. But notice that a structure such as the dog next door, my cat terrifies him is still possible, because in this form CAT is the grammatical subject and DOG is the object (dog is the object in both the preposed position and in the pronominal form him). The difference between the regular active and this left-dislocated construction is a matter of constituent ordering, which is left undecided at this stage of grammatical encoding.

The second component of syntactic processing takes place at the positional level, which operates on the functional-level representation. At this point, serial order is imposed on the utterance. Beginning with the initial constituent, each grammatical function created earlier (e.g., subject, object, modifier) is translated into a linearized constituent. The grammatical encoder retrieves a prestored phrasal frame, which contains slots for all the elements of that phrase – the determiner my and the noun cat, in the current example. Inflectional affixes are represented as an intrinsic part of the frame, so that if the subject were plural, the plural morpheme would already be in place and would therefore not have to be separately retrieved and inserted. Because the language production system is assumed to be incremental (see Section 3.3. for a more thorough discussion of incremental production), the order in which lemmas are 'worked on' determines the overall order of the phrases in the utterance (F. Ferreira, 2000). So if the lemma for DOG were processed before the one for CAT, then the resulting structure might be the left-dislocation form given above or perhaps a topicalized form such as Mary my cat loves. (Although this construction is disfavored in most dialects of English, it can be acceptable given the appropriate context). Thus, positional-level processing determines both the serial order of phrases and the order of elements within any given phrase, and all inflectional processing takes place at this stage of processing as well. For many ordering decisions, the processor simply obeys grammatical constraints such as the requirement that determiners initiate an NP, that adjectives precede nouns, and that verbs precede objects but follow subjects (for English). But because languages give speakers some ordering options, there may be decisions about order that still need to be made, particularly at the within-phrase level. One good example is the sequencing of conjuncts, illustrated in the dog and cat slept soundly. The other order of dog and cat is equally grammatical, and so the choice about how to sequence the conjuncts within the overall NP subject must be based on extra-grammatical considerations (Pinker & Birdsong, 1979; Bock, 1987).

2. Evidence for Two-Stage Models

The evidence for this two-stage architecture separating functional and positional-level processing comes from two sources: speech error analyses and data from experiments designed to shed light on how structure is created. The argument from speech errors is as follows. First, speakers sometimes make semantic substitutions, as in my cat terrifies the boy next door when the girl next door was intended. These errors almost invariably respect a form-class constraint: nouns substitute for nouns, verbs for verbs, and so on. Speakers also sometimes make word exchange errors, illustrated by my boy terrifies the cat next door. The interacting elements in these errors usually come from different phrases, and the words tend to be of the same form class. Semantic substitutions and word exchange errors indicate that there is a level of processing at which grammatical category is relevant and at which the roles for lemmas are decided. In contrast, errors such as phonological substitutions and stranding occur as well, but they have quite different characteristics. In a phonological substitution, a word with a similar sound is incorrectly assembled and made part of the utterance. In stranding errors, content morphemes end up misordered but inflectional material does not, as in I went to get my park trucked (Garrett, 1980). Notice that the morphemes truck and park swapped places, but the suffix –ed is in its correct location. Phonological substitutions and stranding errors indicate that there is a level of processing at which sound and serial order are decided, and stranding suggests that the inflectional morpheme is an intrinsic part of the phrasal frame.

The experimental evidence for distinguishing functional and positional level processing comes from priming studies, both lexical and syntactic. Let us begin with lexical priming. Notice that the two-stage architecture divides lexical processing so that word meanings become available at the functional level, but word phonology only gets generated after (and probably after most positional-level processing takes place as well; F. Ferreira, 1993). This is because the sounds of words are (arguably) not relevant for deciding grammatical functions such as subject and object, but (again, arguably) phonology may help the system decide how to sequence words, as suggested by the finding that, in conjuncts, short words tend to precede longer ones (Cooper & Ross, 1975; Bock, 1987), for example. Experiments in which words are either semantically or phonologically primed have demonstrated that making a lemma available (i.e., semantically priming a concept) causes the constituent containing that lemma to be the subject of the sentence. In contrast, phonological priming has either weak effects or leads to late positioning of the constituent containing the word (Bock, 1987; cf. Cleland & Pickering, 2003). This pattern is typically taken to support a division of labor between functional and positional level processing, because the idea is that only a manipulation, which affects the lemma can influence processes hypothesized to be taking place at the functional level. Of course, this interpretation is somewhat compromised by the finding that phonological priming sometimes leads to late constituent placement, but the effect is much smaller and has been argued to reflect a late stage in production where an utterance is evaluated and then changed if it is judged to be deficient before it is articulated (Levelt, 1989).

Another source of evidence for the two-stage architecture comes from syntactic priming. If a speaker produces or even simply hears an utterance with a particular structural form, he or she is likely to mimic that structure in a subsequent utterance. The classic demonstration (Bock, 1986b) involves both the active/passive and the prepositional/ double-object dative alternations. Speakers will tend to produce a passive sentence after hearing or producing one themselves (Levelt & Kelter, 1982; Schenkein, 1980); the same goes for the prepositional dative (the driver showed the overalls to the mechanic) and the double-object dative construction (the driver showed the mechanic the overalls). (It is not clear that the active can be primed, possibly because of ceiling effects due to its high frequency.) Hartsuiker and Westenberg (2000) discovered using Dutch that a very low-level ordering decision (the sequencing of an auxiliary and a main verb at the end of a sentence) can be primed, leading them to argue for a two-stage model of syntactic processing where a ‘dominance-only’ representation (i.e., one that is not linearized) can prime a representation that is ordered.

Further evidence for the two-stage architecture comes from the way speakers compute agreement relations during sentence production. Consider the fragment the spokesman who defended the actions. If this fragment is the subject of a sentence, then it must agree with the main verb. In English, this agreement process is visible mainly on forms of to be and to have (particularly in the past tense), but in other languages agreement is overt on a wide range of verbs and other words. Carefully designed experiments have revealed that agreement errors occasionally happen, particularly in examples like the spokesman who defended the actions, in which the head noun spokesman is singular but there is another noun in the subject (actions) that is plural (Bock & Eberhard, 1993; Bock & Miller, 1991). Agreement errors turn out to be just as likely in yes/no questions as in declaratives, suggesting that agreement is computed on a representation that specifies dominance but not linear relations (Vigliocco & Nicol, 1998). This argument can be seen in contrast between the helicopter for the flights are safe and Are the helicopter for the flights safe, where the linear positions of the head noun are different but the likelihood of an agreement error is the same. This result suggests that agreement relations are computed from a syntactic representation created before linearization takes place.

Notice that this particular finding is consistent not only with a two-stage view of syntactic processing, but also some version of a transformational/derivational account of grammar, because the linearization process at issue here is the one that moves the verbal material to the front of the sentence to create an interrogative construction. This general idea will be discussed in Section 2.4. when we consider the question whether syntactic representations created during production show evidence of processing attributable to constituent movement. Additional evidence for the idea that hierarchical position but not linear order is critical for computing agreement can be found in a study of complex NPs such as the computer with the program(s) of the experiment(s) (Franck, Vigliocco, & Nicol, 2002). Agreement errors were found to be more likely when the medial noun program was plural compared to the more proximate noun experiment, indicating that position in a hierarchical structure has more effect on agreement than does linear position.

Thus, evidence from speech errors, from syntactic and lexical priming, and from the process of computing subject–verb agreement seem to converge on the idea that syntactic structure is generated in two distinct stages during production. Nevertheless, this architecture has been challenged, and we turn now to evidence that is argued to support a single-stage model of grammatical encoding.

3. Evidence Challenging Two-Stage Models

First, recall the effects of semantic and phonological priming on grammatical form. If it had turned out that only semantic primes could affect the establishment of grammatical relations, then an architecture separating syntactic generation into a stage that uses only lemma information to assign roles such as subject and object, and a separate stage that uses sound to determine linear order, would have been supported. But recall that phonological primes do have a small but significant effect (Bock, 1987). For example, if participants encountered the word trump and then a picture of a truck towing a car, they were likely to say the car is being towed by a truck, because the phonological relationship between trump and truck leads to some type of inhibition. Thus, the effect of a phonological prime appears to be opposite from one that is semantic, but the important point is that according to the classic two-stage architecture, it should have no effect at all. Therefore, it may be argued that this finding undermines two-stage models.

However, there are two problems with this argument. The first was briefly mentioned earlier: It is possible that this effect of the phonological prime occurs not during grammatical encoding but during a stage at which the utterance is checked for overall acceptability (the so-called monitor; see Hartsuiker, Corley, & Martensen, 2005, for a recent discussion of its properties). The second problem with this argument is that the inhibitory effect of phonological primes only challenges the assumptions regarding lexical processing during grammatical encoding – specifically, the idea that lexical retrieval occurs in two stages, with only the second including access of phonology. It is possible that semantic, syntactic, and phonological information about words is all retrieved simultaneously, but that dominance and linear relations are nonetheless computed separately. An important question too is why a phonological prime should be inhibitory rather than facilitatory. Bock (1987) suggested that the effect could be due to lateral inhibition among phonological competitors, but some studies of lexical processing have shown that phonological primes facilitate processing (Grainger & Ferrand,1996; Tanenhaus, Flanigan, & Seidenberg, 1980). The

Bock (1987) finding clearly should be pursued further, and indeed, it has not yet even been replicated.

The second finding that has been argued to undermine the two-stage model of syntactic processing concerns priming in the dative structure (Pickering, Branigan, & McLean, 2002). Consider once again the prepositional-dative, illustrated in the driver showed the overalls with the stains to the mechanic. Another grammatical alternative is the shifted form in which the prepositional phrase (PP) precedes the object (the driver showed to the mechanic the overalls with the stains), a structure that is more likely to be generated the longer and heavier object (Wasow, 1997). Shifted and non-shifted prepositional datives share the same hierarchical or dominance relations but differ in how the NP and the PP are ordered. Thus, on a two-stage view in which dominance relations are computed separately, the shifted version should prime the non-shifted version. However, such priming does not occur. Based on these results, Pickering et al. (2002) argued for a single-stage model in which dominance and linear relations are computed simultaneously. But the results could be attributed to the peculiarities of the shifted dative form, which is not only fairly rare (even in the Pickering et al. experiments in which measures were taken to elicit them) but also seems to require fairly strict discourse conditions to be felicitous (Hawkins, 1994). These properties of the shifted prepositional dative might compromise its ability to prime any other construction. It would be very useful to see whether this result can be found using a less marked structure. Exploring this possibility might require consideration of languages that allow more flexibility in constituent ordering than English does.

4. Do Syntactic Structures Contain Evidence of Constituent Movement?

Perhaps the most distinctive characteristic of generative grammar compared to other approaches to syntax is its assumption that syntactic structures are generated by movement. Anyone who has taken even an undergraduate course in cognitive psychology knows that in the earliest versions of this theory, noncanonical structures such as passives were created by rearranging the basic active structure (Chomsky, 1965). Somewhat less well known is the transition to the Government and Binding (GB) theory, which assumed that syntactic representations contain evidence of movement. For instance, a passive such as the dog next doori was bitten ti by my cat requires movement of the NP the dog next door from the post-verbal position to the subject position, but the starting position of the phrase is marked in the representation with a trace (indicated with the t). The trace allows the phrase to be interpreted as the object of the verb bite even though it is no longer in object position in the surface structure. The same holds for structures such as wh-questions and relative clauses: A sentence such as which dogi did my cat bite ti ? is created by moving the wh-NP (which dog) to the top of the tree, again leaving behind a trace so that dog can be interpreted as the object of bite. It has been common in the psycholinguistic literature to refer to traces as gaps and to moved constituents as fillers (J. D. Fodor, 1978, 1989, 1991), and so we will follow this convention for the rest of our chapter.

The question we now turn to is, do the syntactic structures that people create when they talk contain any evidence of constituent movement? It is widely believed that they do not. For example, it has been argued that one way to conceptualize the two-stage architecture for grammatical encoding is to assume that the first stage creates a 'deep-structure' representation and that the second creates a 'surface structure' representation. It is important to note, however, that the concept of a 'deep structure' has really not been part of generative grammar for the last 25 years, and so it would be surprising to find any evidence for it in language production. And, not unexpectedly, we do not.

Bock, Loebell, and Morey (1992) used the syntactic priming technique to distinguish between the direct and mediated (first a deep structure is computed, then a surface structure) approaches to syntactic generation. Participants heard sentences and then repeated them, and then they had to describe an unrelated picture of a simple transitive event. The critical feature of the study was that if the pictures were described in the active voice, the subject would be inanimate and the object animate (e.g., the clock woke up the boy). The prime for the picture description (i.e., the heard and repeated sentence) was either active or passive, and it had either an animate or an inanimate subject. The assumptions behind the design were that both structural form and animate placement would be mimicked in the picture descriptions. Bock et al. reasoned that if the 'deep structure then surface structure' hypothesis is right, then both an active with an inanimate object AND a passive with an inanimate surface subject would prime the active picture description, because the passive actually has an inanimate object at deep structure. However, this pattern was not observed; instead it was the surface placement of the animate entity that determined the degree of animacy priming. Hence, Bock et al. concluded that the direct mapping account is correct.

Bock, Loebell, and Morey (1992) used the syntactic priming technique to distinguish between the direct and mediated (first a deep structure is computed, then a surface structure) approaches to syntactic generation. Participants heard sentences and then repeated them, and then they had to describe an unrelated picture of a simple transitive event. The critical feature of the study was that if the pictures were described in the active voice, the subject would be inanimate and the object animate (e.g., the clock woke up the boy). The prime for the picture description (i.e., the heard and repeated sentence) was either active or passive, and it had either an animate or an inanimate subject. The assumptions behind the design were that both structural form and animate placement would be mimicked in the picture descriptions. Bock et al. reasoned that if the ‘deep structure then surface structure’ hypothesis is right, then both an active with an inanimate object AND a passive with an inanimate surface subject would prime the active picture description, because the passive actually has an inanimate object at deep structure. However, this pattern was not observed; instead it was the surface placement of the animate entity that determined the degree of animacy priming. Hence, Bock et al. concluded that the direct mapping account is correct.

Recent evidence from errors of subject-verb agreement also support the idea that traces are mentally represented during grammatical encoding (Franck, Lassi, Frauenfelder, & Rizzi, in press). An experiment designed to elicit such errors from French speakers showed that displaced direct objects in a cleft construction (It’s the deputy that the senators welcome t) determine whether errors of agreement occur, even though the object does not intervene between the head and the verb in the surface word sequence. Franck et al. argue that their results can only be accounted for if we assume not just a single transformational process that turns a set of lexical items into a surface structure, but rather a grammar consistent with the Minimalist Program

(Chomsky, 1995) in which syntactic structures are generated through a series of operations termed MERGE, MOVE, and AGREE (only the latter two are relevant for our purposes; MERGE simply refers to the combining of lexical items). According to Minimalism, elements in a syntactic structure cyclically move up the tree until they are in the appropriate position to allow an agreement relation to be checked. Each movement, even intermediate ones, leaves behind a trace. Franck et al. argue that the pattern of agreement errors they observe in French and Italian can only be explained if it is assumed that gaps of intermediate movement interfere with the process of computing agreement. These results are some of the strongest evidence to date that gaps are generated as part of the normal process of creating a syntactic structure for a sentence.

What should be said, then, about the work suggesting that gaps do not exist? At this point, the most unbiased assessment of the state of our knowledge is that this entire issue needs to be examined in much more detail. In fact, it is worth noting that no experiment has ever been conducted to test directly whether gaps are psychologically real in language production (although the Franck et al. paper comes close). One potentially useful observation about our current state of knowledge is that the evidence for and against traces comes from different sources – the evidence against the reality of gaps is based largely on results from syntactic priming, and the evidence for them has come from studies of subject-verb agreement as well as the process of translating syntactic structures to prosodic constituents. Of course, the traces of wh- and NP-movement are not the only types of empty categories that have been proposed in generative grammar; another important phonetically null category results from ellipsis, as in Mary can tie her shoes and Natalie can too. Comprehension studies have shown that people reconstruct the missing material, eventually obtaining the interpretation that Natalie can tie her own shoes (perhaps by first entertaining but then rejecting the so-called ‘strict’ reading on which Natalie ties Mary’s shoes; Shapiro, Hestvik, Lesan, & Garcia, 2003). The production of ellipsis has not been studied at all, so the extent to which the omitted or deleted material is mentally represented is not known. Consider our shoe-tying example. It is clear that lexical forms corresponding to the second verb phrase tie her shoes are not retrieved. But at the message level, the speaker almost certainly generates the idea that Natalie is capable of tying her own shoes. The question those interested in grammatical encoding might ask is, what about the levels in-between? Is a VP generated in the second clause and then left unpronounced because the lemmas do not point to any word-forms, as Levelt (1989) suggests? Another intriguing question is what leads speakers to choose one type of ellipsis over another; for example, an alternative to the form above is Mary can tie her shoes and so can Natalie. Clearly, many important issues concerning the grammatical encoding of empty categories remain to be even formulated in the field of language production.

5. Are Syntactic Structures Lexically Anchored?

Another question that has been of interest both in formal linguistics and in psycholinguistics is whether syntactic structures are linked to words – specifically, lemmas. The classic work by Levelt (1989) argued for lexical generation of syntax. In contrast, the BL model of production assumes a non-lexical view of syntactic structure. In BL, trees are conceptualized as 'control hierarchies,' which contain no lexical content but instead coordinate the insertion of lexical material which is retrieved and assembled separately (BL, pp. 947–948). This conceptualization is in part based on Bock's earlier findings suggesting that lexical overlap does not enhance syntactic priming (Bock, 1989; Bock & Loebell, 1990). For example, the amount of priming for a sentence such as the girl handed the paintbrush to the man is the same given a prime like the secretary baked a cake for her boss and the secretary gave a cake to her boss, even though in the latter case the PPs share the same prepositional head (to).

However, more recent work using the syntactic priming paradigm suggests that verb identity does increase the magnitude of priming (Pickering & Branigan, 1998, 1999; Cleland & Pickering, 2003). One motivation for examining this issue in careful detail is that many formal theories of syntactic structure assume that words and the syntactic environments in which they may occur are lexically linked. In the earliest versions of transformational grammar, for example, verbs specified the syntactic environments in which they could occur via their subcategorization frames (Chomsky, 1965). A verb such as put would be represented as requiring an object and a locative PP. The theory of GB essentially dissolved phrase structure rules altogether in favor of lexical storage of constituent structure, so that all words were represented in the lexicon with their associated arguments (Chomsky, 1981; Stowell, 1981). Retrieval of a word would then automatically bring along its associated structures. This elimination of phrase structure rules was a logical extension of X-bar theory (Jackendoff, 1977), which described a universal format for all phrases regardless of their type. Other models of syntax such as LFG (Bresnan & Kaplan, 1985), Categorical Grammar (Moorgat, 1988; Steedman, 2000), and Generalized Phrase Structure Grammar (Gazdar, Klein, Pullum, & Sag, 1985) also connect words and their syntactic environments. The same trend is evident in Tree-Adjoining Grammar (TAG) (Joshi, Levy, & Takahashi, 1975; Joshi, 1985). In TAG, the primitive objects of the grammar are treelets, which consist of a word (a lexical head such as a verb) and the arguments the head licenses.

Given these theoretical perspectives on the representation of words and syntactic structures, and in particular verbs and arguments, it makes sense to expect that syntactic priming would be greater when the main verb in the prime and target sentences overlap. This issue was investigated in a study designed to assess whether priming would be observed in simple dialogue situations (Branigan, Pickering, & Cleland, 2000), and which manipulated verb identity. A confederate and a genuine experimental subject described pictures to one another, and it was found that the naïve participant tended to use the same construction as the confederate. In addition, priming was greater when the verb in the prime and target sentences was the same – the effect was about twice as large. The model of syntactic generation offered by Pickering and Branigan (1998) assumes that words such as verbs are linked to the phrases with which they may combine (termed combinatory nodes). Cleland and Pickering (2003) demonstrated remarkably similar priming effects for noun phrase structure, including enhancement when the head of the noun phrase (the noun) was shared, indicating that this form of representation and the process for creating structures are similar both for clauses and phrases.

These results, then, tell us that the syntactic representations used for language production are ones in which structures may be generated directly from lemmas rather than through the accessing of contentless phrasal templates (Levelt, 1989). F. Ferreira (2000) presented a model for human language production, which uses TAG as the database for creating structures through lemmas (see also F. Ferreira, Lau, & Bailey, 2004). All heads, including verbs, nouns, prepositions, and adjectives, are represented with the arguments that they license. These elementary trees consisting of a head and its licensed arguments are combined to form utterances (see F. Ferreira, 2000 for a description of the operations that combine elementary trees). The model proposed by Pickering and his colleagues is somewhat different, but it shares the basic insight that words and syntactic structures are representationally linked. Thus, although it may be possible for grammatical encoding to take place using control structures that have no lexical content, perhaps via extraction of some type of general schema for forming particular construction types, in general it appears that the syntactic structures used for production are lexically anchored.

B. Processing Resources for Grammatical Encoding

Thus far, we have considered mainly representational issues, focusing particularly on the properties of syntactic structures and the format in which syntactic information is stored. The question we turn to in this section is how these structures are formed, and in particular, how the process of grammatical encoding draws on processing resources, and how computational load is managed. This will lead us to consider the degree to which grammatical encoding is incremental.

1. Is Grammatical Encoding Automatic?

Talking generally feels effortless, but even the most fluent speakers occasionally experience some difficulty formulating their utterances. The classic work of GoldmanEisler (1968) demonstrated that almost half of most people's speaking time is devoted to pauses and disfluencies such as um and er. Ford (1982) measured spontaneous speech and observed that about 20% of all clauses are preceded by a pause of about one second in duration. This finding suggests that the process of creating syntactic structure is resource-demanding, but the result is not definitive; clauses are both major syntactic and semantic junctures, and it might be that grammatical encoding is automatic but semantic processing requires planning and can therefore be resource demanding. This view was articulated by Levelt (1989). He adopts Kempen and Hoenkamp's (1982, 1987) model of grammatical encoding which assumes that syntactic procedures are modular, and thus have the characteristics that J. A. Fodor (1983) views as typical of cognitive modules:

grammatical encoding processes consult a proprietary vocabulary, they operate whenever they recognize their standard input, and they operate automatically. Thus, it is to message level planning that resources are devoted during production; syntactic decisions are made automatically, and measures of processing load such as reaction time and pausing will not be affected by the complexity of syntactic operations.

Work conducted since the publication of Levelt's book, however, does appear to suggest that syntactic planning demands computational resources, especially as structural complexity increases. F. Ferreira (1991) had people memorize declarative sentences, which they then had to produce upon receipt of a visual cue; latency to begin speaking was measured. The variable that was manipulated was the syntactic complexity of the sentential subject. It was either short (the river) or long, and in the long conditions, it was either of low (the large and raging river), medium (the river near their city), or high (the river that stopped flooding) syntactic complexity. Complexity was defined in terms of a node count, so that the more syntactic nodes the subject needed in its representation, the greater its complexity. Ferreira found that as complexity increased, so did production latencies. Interestingly, memorization times were not affected by this variable, suggesting that the effect was particular to the task of speaking. In a second experiment, she orthogonally varied the syntactic complexity of the subject and object in subject-verb-object sentences. Once again, latencies to begin speaking increased with the complexity of the subject, and the object's characteristics had no effect. However, when both the subject and the object were syntactically complex, speakers tended to pause within the sentence, and the preferred pause location was the subject-verb phrase boundary.

Because speakers were not required to generate any of the sentences' content, these effects of syntactic complexity cannot be attributed to any semantic complexity that might be correlated with the syntactic manipulation. Moreover, as speakers had not chosen the syntactic forms themselves either, the effect cannot be attributed to the need to make syntactic decisions. Instead, it appears that simply saying a sentence with a complex structure takes up processing resources. The results also demonstrate that if both the subject and object are complex, speakers divide the utterance into two processing units, one consisting of the subject, and the other consisting of the verb phrase.

Notice that this division into processing units respects the syntactic structure of the sentence, and indeed Ferreira observed almost no cases in which participants paused after the main verb rather than before it (see also F. Ferreira, 1993 for further discussion of these issues). This finding that the processing units are syntactic constituents is consistent with the assumption that the difficulty in processing is localized to the syntactic level. In another study, Smith and Wheeldon (2001) tested whether speakers plan their utterances before saying them. Participants were asked to describe pictures, and they were primed with sentences such as The spoon and the car move up. When participants uttered sentences that were syntactically similar to the prime sentence, a reliable 50 ms advantage to begin speaking was observed. Smith and Wheeldon also tested the scope of this effect and found that it held only for the first phrase of an utterance, consistent with F. Ferreira's (1991) second experiment demonstrating that only the complexity of the subject affects utterance initiation times. The advantage that Smith and Wheeldon found for sentences that had been syntactically primed suggests that we should revisit the phenomenon of syntactic priming in light of this question concerning processing resources. Recall that speakers are more likely to produce a particular construction when they have just heard it or produced it themselves. Based on the finding that syntactic priming is particularly robust in dialogue, Pickering and colleagues have suggested that syntactic priming is used by the production system as a tool for "reducing the load associated with syntactic processing" (Pickering & Branigan, 1999, p. 136). What this idea assumes, of course, is that syntactic generation is a resource-demanding process, so much so that speakers try to find ways of managing and reducing the computational burden.

It should be noted that studies predating the recent psycholinguistic era also demonstrate that syntactic processing is computationally demanding. Johnson (1966) compared the generation of sentences such as The person who jumped over there is good and The person over there who jumped is good. Because the structure of the second sentence is right-branching, it is less complex according to the Yngve (1960) complexity metric, and so Johnson predicted it would take less time to initiate. This prediction was confirmed. Second, although Rochester and Gill (1973) did not find any effects of what they termed "syntactic complexity" on speech hesitations and disruptions, they did find that such disruptions in speech varied along with the type of nominal modifier people produced. Specifically, speakers were more likely to show speech disruptions before a noun phrase complement (e.g., “The fact that the woman was aggressive threatened the professors”) than before a relative clause (e.g., “The book that was written by Millet was lauded by all”). Goldman-Eisler (1968), who like Rochester and Gill (1973) failed to find effects of syntactic complexity on hesitation, also found hesitation differences before different types of syntactic forms. If disruptions in speech are a measure of mental load, and more disruptions occur before one particular ordering of words than another, then one structure must have required the use of more mental resources than the other. We turn next to a more detailed consideration of this question concerning the inherent difficulty of certain forms.

2. Are Some Constructions Difficult to Generate?

In this section, we ask a question that has received surprisingly little attention from experimental psycholinguists. Are some syntactic constructions inherently difficult to produce, or does difficulty arise only when a structure must be generated in an infelicitous discourse context? To see what is the issue here, consider the passive construction, which is often viewed as more complex than the active, and is certainly more difficult to understand (F. Ferreira, 2003). The passive may be harder to produce than the active because it has a noncanonical structure, because it is less frequent, or because it is more complex, in the sense of requiring more syntactic nodes in its phrase-structure representation. Alternatively, it has been argued that the passive may be the *right* construction for particular discourse situations. For example, Tomlin (1983) observed that passives are very common in hockey broadcasts, because what the commentator tries to do is make the player in possession of the puck the subject of the sentence. If that player is affected in some way (e.g., gets checked), then the sentence form the commentator will use is a passive (Gretzky was checked by his opponent), because that is the form that allows the topic to be maintained as subject, even when it is not an agent. But Tomlin’s study did not examine whether passives are harder to produce even when they are licensed by the discourse. To answer this question, it is necessary to measure processing load rather than just frequency of occurrence.

Tannenbaum and Williams (1968) conducted one relevant study. Speakers first read a story that was either about trains, cars, or a topic that was relatively neutral. They then saw a picture of a train hitting a car, and their task was to produce either an active or a passive sentence (cued by a letter superimposed on the picture). They found that latencies to produce the active were fastest in the subject-focus condition and about equally long in the object-focus and neutral conditions. Passives were produced fastest in the object-focus condition, next fastest in the neutral condition, and most slowly in the subject-focus condition. This finding would appear to suggest that as long as a construction occurs in the appropriate context, it is easy to produce. However, a closer examination of their data suggest a different conclusion. Although this pattern was observed, it was also found that actives in the "wrong" discourse were produced as quickly as passives in the "right" discourse; indeed, in no condition were passives initiated faster than actives. The picture that emerges, then, is that noncanonical structures can be inherently hard to say, even in proper contexts. A construction that is rare or that is syntactically complex (or both, as these two characteristics tend to co-occur) requires a specific sort of context but is still difficult to generate, perhaps because more syntactic nodes take more processing resources to create, or because the production system has less experience generating forms such as the passive. This finding is compatible with a study that investigated whether certain verbs license passives more easily than others, as might be expected on a lexicalist view of grammatical encoding (F. Ferreira, 1994). Participants were asked to generate sentences out of three visually presented words – e.g., LAYOFFS MANAGER WORRIED. The verb either had a theme-experiencer argument structure (as in worried) or a more conventional agent-patient structure (LAYOFFS MANAGER ORDERED). Speakers produced passives more often when the verb was theme-experiencer, which was predicted based on the idea that speakers attempt to place the more prominent thematic role in the subject position of the sentence, and experiencers are more prominent than themes (Grimshaw, 1990). Nevertheless, passives took longer to formulate than actives, suggesting that even though certain lexical conditions might license them, they still seem to take more time to grammatically encode.

Clearly, however, this issue needs to be examined in more detail, particularly now that there is such intense interest in the idea that the frequency of exposure to a syntactic construction affects how easily it can be comprehended (MacDonald, Pearlmutter, & Seidenberg, 1994; Mitchell &

Cuetos, 1991). As Race and MacDonald (2003) have pointed out, these distributional patterns come from speakers – they reflect the choices speakers make in different circumstances. The Race and MacDonald approach to processing assumes that comprehension and production must be examined together, and they predict that the forms that are hard for people to produce are also the ones that are hard to understand. This parity is based on speakers' tendency to avoid difficult structures, thus creating distributional patterns. But this interesting research program is predicated on the idea that some forms are inherently difficult to produce – for example, an object relative such as The story the quiet boy read was long, which does not contain the relative pronoun that – and that is why it is less likely to be said. The empirical question that arises is whether these sentences are harder to say when they are generated, or whether the discourse conditions, which obtained at the time the sentence was said in fact made the structure easy to encode and articulate.

3. Incremental Production

Incremental production may be viewed as a way to reduce the processing resources required for production. The idea is that at particular points in time certain concepts may be more available to a speaker than others, and the grammatical encoder tends to begin with those accessible lemmas. Incrementality is viewed as optimizing the use of processing resources, because it allows the system to begin with the 'easy bits', so to speak, and to deal with the more difficult portions of the utterance during articulation (F. Ferreira & Henderson, 1998). Of course, as a reviewer of an earlier version of this chapter pointed out, incrementality might create a situation in which an accessible constituent forces a syntactic structure that is computationally demanding (e.g., the passive). But the reason incrementality will generally still lead to efficient processing is that the difficulty of making a passive can be 'spread out' over the entire utterance rather than being localized entirely to the point of its initiation. As a result, there need not be any hesitation or disfluency before utterance production, and the demands of managing the rest of the structure can be distributed over the remaining constituents, with planning going on in parallel with articulation.

Recent work suggests that the degree to which the system is incremental is under strategic control, as would be expected if incrementality is a way for the production system to manage its resources (F. Ferreira & Swets, 2002). Participants were asked to calculate the answers to arithmetic problems and to provide the answer in the form of a sentence (The answer is 58). The problems always included at least one two-digit addend (e.g., 53+5), so participants were unlikely to be able to retrieve the sum. In the first experiment, participants were allowed to begin to speak whenever they felt ready, and the data provided no evidence for incremental production. Initiation times were longer the more difficult the entire problem, but durations were unaffected. This pattern indicates that the entire utterance was planned before articulation. In the second experiment, speakers were required to begin to speak before a deadline (indicated by a punishment ‘beep’). This manipulation dramatically reduced initiation times overall, from over 2 s in the first experiment to about 700 ms in the experiment with the deadline (interestingly, accuracy was not compromised). Nevertheless, initiation times still reflected the difficulty of computing the sum. At the same time, the duration of the earlier part of the utterance was also affected by problem difficulty, suggesting that speakers postponed some planning of the sum until they were actually speaking. This study suggests that the degree to which the system is incremental depends on the speaker’s strategy. If a premium is placed on beginning to speak quickly, then the production system does indeed become more incremental; but if speakers have the opportunity to plan, they seem to prefer to do so. Moreover, the system engages in some planning even under conditions most conducive to incremental production – that is, when there is a premium on initiating speech quickly. On some views of incrementality, constructions are chosen indirectly; they emerge from the speaker’s attempt to place a highly accessible concept in the most prominent syntactic position. If that concept happens to be a theme or patient, in a relatively fixed word order language such as English a passive structure will need to be produced to accommodate a thematic patient in subject position. Under this view of human sentence production, the lemmas associated with the most accessible concepts automatically grab the earliest positions in utterances. However, Bock (1986b) questioned this radical version of incrementality: “Typically, speakers do not simply produce words in the order in which they come to mind …. Rather, the syntactic forms of sentences seem to be changed so as to accommodate word order variations without altering the intended meaning” (p. 359). But a radically incremental model is assumed, for example, by van Nice and Dietrich (2003), who interpret their German data as supporting the view that "the firstconceptualized referent will continue onward as the first-lexicalized and, ultimately, as the first in word order" (pp. 829). This view, they point out, is also held by Kempen and Hoenkamp (1987) as well as de Smedt (1996).

Christianson and F. Ferreira (2005) attempted to resolve this controversy by examining production in Odawa, an Algonquin language which allows constituents to be ordered freely (i.e., any arrangement of subject, verb, and object is grammatically licensed). Speakers were asked questions about a pictured event. The questions topicalized either the agent, the patient, or neither (this latter question was simply, What happened?). Even though Odawa speakers have access to any word order arrangement of subject, verb, and object, their descriptions were similar to those observed for English speakers. Given the no-topic and agent-topicalizing questions, actives were the forms most commonly produced; in the patient-topicalizing question condition, passives were preferred. Thus, even though speakers of Odawa could have produced active sentences with the patient in the first position (i.e., OSV or OVS sentences) when the patient was topicalized, they in fact chose to produce passives, which not only are about as rare in Odawa as they are in English, but also require the omission of the agent argument altogether (because passives in Odawa do not permit any type of by-phrase). Thus, a highly available constituent primes a particular syntactic form, and if that constituent is a patient, the form that will be generated is a passive. These findings are inconsistent with extreme versions of incremental production, and instead support V. Ferreira and Dell (2000), who argued that the lexically driven picture of production – in which the most accessible lexical item wins a figurative "race" out of the mouth – might not be sufficient to accurately describe their results. Instead, they proposed that speakers choose a syntactic structure without necessarily first deciding between alternative lexical items. The structure, then, is what is really primed by pictures, sentences, and questions. On this view, incrementality applies to the filling of available NP nodes in the primed structure (F. Ferreira, 2000). Incrementality encourages the selection of a syntactic structure that allows accessible material to be mentioned sooner (V. Ferreira & Dell refer to this as 'lexical-syntactic interactionism').

Another way to think about incrementality in production is to ask what sorts of planning units the system uses. The most extreme versions of incrementality assume that there is little or no look-ahead, and so predict that planning units will be essentially non-existent (i.e., utterances are planned more or less word-by-word). On the other hand, non-incremental views assume that the system does engage in look-ahead over some multi-word domain. Clauses have been classically assumed to serve as planning units for grammatical encoding (Boomer, 1965; Ford & Holmes, 1978; Garrett, 1975; Lashley, 1951). The idea is that the system organizes an entire clause (i.e., a verb and its arguments) before engaging in any phonological encoding. One way that this issue has been addressed is by examining how speakers compute grammatical agreement between complex subjects and inflected verbs. A variety of studies have demonstrated that a distractor noun in the subject can be an attractive lure for agreement, especially if it is plural (Bock & Eberhard, 1993; Bock & Miller, 1991). Bock and Cutting (1992) used this phenomenon to determine whether the unit of grammatical encoding is the clause. They varied whether the constituent that intervened between the head noun of the subject and the main verb was a PP modifier or a relative clause (e.g., the editor of the history books versus the editor who rejected the books). They reasoned that if clauses are planning units for grammatical encoding, then agreement errors (plural inflections on a form of to be) should be less common in the relative clause condition. This prediction follows because the relative clause would be planned separately, and thus the head noun and the verb would be more closely linked during processing. This prediction was confirmed: Agreement errors were more common when a relative clause came between the head noun of the subject and the main verb, consistent with the classic idea that the unit of syntactic planning is the clause. This finding is inconsistent with radical incrementality or any type of production system which generates utterances on a word-by-word basis, but it can be reconciled with more limited incrementality (Christianson & F. Ferreira, in press; F. Ferreira, 2000).

Finally, there is evidence that the production system operates more efficiently when it has syntactic options that allow potentially different states of activation to be taken into account during grammatical encoding. V. Ferreira (1996) compared the production of sentences headed by a verb such as give, which alternates between a double-object and a prepositional dative form, and verbs such as donate, which only allow the prepositional dative (e.g., *The widow donated the library her entire collection). He found that sentences with syntactically flexible verbs such as give were generated more quickly and more fluently than sentences with more restrictive verbs. He argued that flexibility allows the system to accommodate lemmas' potentially different states of activation over time. For example, if a speaker has said The widow gave and then finds that direct object hard to retrieve, he or she can continue processing by working on the indirect object instead, because the verb give permits this flexibility. Thus, one benefit of syntactic freedom of choice is that it enhances the efficiency of language because the verb give permits this flexibility. Thus, one benefit of syntactic freedom of choice is that it enhances the efficiency of language production. In the next section, we focus specifically on the issue of how speakers make syntactic choices.

C. Syntactic Choice

1. Choice of Syntactic Construction

As already mentioned, work by Carroll (1958), Bock (1986a, b), Bock and Warren (1985), and others has shown that, in English, when a noun phrase is made accessible by showing someone a picture of a semantically related item, asking a focusing question, or establishing a context, speakers tend to begin their sentences with that primed NP. Bock and Warren's (1985) work on the production of passives and dative structures in English indicates that the most accessible entity claims not only an early position in the string, but also the most prominent syntactic function (i.e., subject or non-oblique dative in ditransitive structures). A similar finding is that passives tend to occur with theme-experiencer verbs, because the passive allows the experiencer to be placed in subject position (F. Ferreira, 1994). This effect is larger when the experiencer is human and the theme is not, indicating that an animacy contrast perhaps helped to distinguish the conceptual prominence of the two entities even more than just their thematic role status. Spanish speakers also are sensitive to accessibility when they choose syntactic constructions (Prat-Sala & Branigan, 2000). Spanish syntax includes a dislocated active structure (OSV, along the lines of 'Cheese I love to eat'), which allows the effects of inherent accessibility (animacy) to be distinguished from those of derived accessibility (discourse prominence). Spanish speakers tended to place the more salient entities in higher syntactic positions, making use of both passives and the dislocated active structure. In general, then, syntactic forms are chosen to allow speakers to line up conceptual and syntactic prominence.

2. Inclusion of Optional Functional Elements

Thus far, we have considered how speakers decide on a syntactic form – active versus passive, double-object versus prepositional dative, and so on. Now we ask a slightly different question: How do speakers decide whether to include an optional function word such as the complementizer that in a sentence like The weary traveler claimed (that) his luggage had been stolen? If the complementizer is omitted, an ambiguity about the status of the noun phrase the luggage is created for the comprehender: the luggage could be either the direct object of claimed or the subject of a complement clause. The presence of the complementizer essentially disambiguates the structure, making it clear that the noun phrase is a subject. (It is possible for that to be a determiner, as in that luggage, not your luggage, but Roland, Elman, and V. Ferreira (in press) have demonstrated that post-verbally, the word that is almost always a complementizer, and the parser is likely tuned to this distributional information.) If speakers attempt to produce utterances that are easy for their listeners to understand, one might predict that, the greater the chance of a misinterpretation, the greater the likelihood that speakers will include the complementizer. For example, if the verb preceding the ambiguous noun phrase subcategorized for only clausal complements, the that is unnecessary, and so it might be omitted; but if the verb takes both direct objects and clausal complements, the that would help the listener avoid making a parsing error.

The evidence suggests that speakers' needs motivate complementizer inclusion. This has been shown in a variety of experiments by V. Ferreira (2003; V. Ferreira & Dell, 2000), which demonstrate that that is more likely to be included in complement and relative clause structures when the speaker is having difficulty retrieving the word that would follow that. Two different mechanisms can be proposed to account for this relationship: Alleviation and Signaling (Jaeger, 2005). According to the Alleviation hypothesis, speakers include that to give themselves time to plan (Race & MacDonald, 2003), making the complementizer essentially like a filler term such as uh. The alternative hypothesis, Signaling, assumes that the complementizer is a signal or at least a symptom of upcoming difficulty. The two hypotheses make opposite predictions about the distribution of complementizers and filler disfluencies. If Alleviation is right, then the presence of a complementizer should reduce the likelihood of a filler. If Signaling is correct, then thats and fillers should be positively correlated. Jaeger (2005) and V. Ferreira and Firato (2002) found results consistent with the second pattern, which supports the Signaling hypothesis. It is important to note, however, that the data are compatible with the idea that complementizer inclusion is merely a symptom of difficulty – that is, the same factors that lead to disfluencies lead to the inclusion of a complementizer as well.

This pattern has emerged in other studies as well-speakers in dialogue tasks fail to make use of either optional words or disambiguating prosody to avoid ambiguity (Allbritton, McKoon, & Ratcliff, 1996; Arnold, Wasow, Asudeh, & Alrenga, 2004; Kraljic & Brennan, 2005). One exception is a recent study reported by Haywood, Pickering, and Branigan (2005), who found that speakers did provide more disambiguating that's when they were describing objects to a conversational partner. However, as V. Ferreira, Slevc, and Rogers (2005) argue, the effect may be due to the visual properties of the situation the interlocutors were presented with. The situation which led to ambiguity in the Haywood et al. study was one in which there was more than one object of the same type, thus inviting the use of a disambiguating modifier (e.g., the penguin THAT'S in the cup on the star), and it is in these situations that the word that tended to be included. Ferreira et al. argue that perhaps speakers were simply sensitive to the existence of more than one token of the same type and in those cases produced more explicit utterances. At the same time, it must be acknowledged that even if this interpretation is correct, it still appears that Haywood et al. have indeed observed the altruistic rather than egocentric use of optional functional elements. Moreover, as highly skilled speakers are probably better able to avoid ambiguity than those who are less practiced, it is clear that some mechanisms must exist to allow speakers to monitor their speech and include optional elements in just those situations when they might be helpful to listeners. What we do not know is how this process, which is potentially quite resource-demanding, is coordinated with the other tasks performed by the production system.

D. Syntax and Prosody in Production

It is clear that utterance stress and timing have something to do with a sentence's syntactic structure. Phonologists have debated whether the correct characterization of these effects appeals directly to syntactic constituents, or instead makes reference to prosodic entities such as phonological words, phrases, and intonational phrases. According to the syntactic view (Cooper & Paccia-Cooper, 1980; Odden, 1990; Selkirk, 1984; Wagner, 2005), the amount of lengthening and stress assigned to a given word can be directly related to syntax. For example, the more syntactic right brackets that terminate on a word, the longer and more stressed it will tend to be. On the prosodic constituency view, syntax is used to create prosodic constituents, but then it is features of prosodic constituency that determine timing and stress (Gee & Grosjean, 1983; Inkelas & Zec, 1990; Levelt, 1989; Selkirk, 1986). Disentangling these two approaches to rhythm can be challenging, because prosodic and syntactic constituency are highly correlated (F. Ferreira, 1993), but one important theoretical difference between them is that prosodic structure is generally viewed as flatter and less articulated than syntactic structure, because prosodic constituency is generally thought not to permit recursion (Selkirk, 1986; cf. Gee & Grosjean, 1983; Ladd, 1986; Wagner, 2005). The idea is that, in syntax, a clause may have another clause inside it (for example), but in prosody, such self-embedding is forbidden. As a result, prosodic structures are flatter than syntactic ones, allowing prosody to serve as an interface between hierarchical and recursive syntactic/semantic representations and the sequential speech channel through which articulation must take place. Another important difference between the two types of structures is that prosodic representations pay attention to the distinction between function and content words. Therefore, a phrase consisting of just a pronoun, for instance, would typically not behave the same way as a full lexical NP.

Intonation is related to syntax too, but again, it has long been known that the intonational phrasing of a sentence may not be isomorphic to its syntactic constituency. One famous example is This is the cat that chased the rat that swallowed the cheese…, which tends to be phrased as (this is the cat) (that chased the rat) (that swallowed the cheese), even though the major syntactic boundary is between is and the cat. Other more realistic examples include (Mary left)(after the party) and (Mary gave the book) (to her brother who lives in Ohio). In both these cases, the major intonational boundary comes not between the subject and verb phrase, but after the verb. To account for these cases, Selkirk (1984) proposed the Sense Unit Condition, which states that the constituents making up a single intonational phrase must be in either a head-argument or head-modifier relation. The Sense Unit Condition rules out apparently malformed examples such as (Ten mathematicians) (in ten derive a lemma) (Steedman, 2000), because in ten is a modifier of mathematicians and is not a head, argument, or modifier of derive a lemma. Steedman (2000) argues that a grammar such as Combinatory Categorial Grammar, which allows a wider range of syntactic constituents than other approaches captures these sorts of facts and eliminates the need for a separate and stipulative Sense Unit Condition. The important point for our purposes, however, is that even though the intonational phrasing of a sentence might ultimately deviate from its syntactic structure, the well-formedness of the intonational phrasing appeals to syntactic concepts such as head, modifier, and argument.

So far we have considered only aspects of prosody that can be directly related to linguistic structures, either prosodic or syntactic. But syntax affects the sound properties of a sentence in another way, which we will roughly characterize as having to do with performance effects. For example, hesitations and pauses due to planning difficulty tend to cluster at clause boundaries (Ford, 1982; Goldman-Eisler, 1968). In addition, it has been argued that the most probable location for pauses and prosodic breaks can be predicted from algorithms, which assume that break points are jointly determined by the complexity of material to the left and to the right of the boundary (Gee & Grosjean, 1983; Watson & Gibson, 2004). F. Ferreira (1991) demonstrated that the syntactic complexity of upcoming material affected pause duration, and she argued that the effects were due to the difficulty of planning upcoming material. Thus, the sound pattern of a sentence has at least two possibly distinct sources: One is the syntactic and prosodic representation which might mandate breaks in particular locations, and the other is the speaker's need for more time to plan upcoming material. An important question for future research on the syntax-prosody interface in language production is whether these two sources are indeed distinct, or whether prosody and performance phenomena can be reduced to the same underlying causes. Another critical issue is whether it is necessary to postulate a distinct level of prosodic constituency to account for phenomena related to rhythm and intonational phrasing, or whether syntactic structure is sufficient to explain prosodic patterns in spoken sentences. One limitation of work that has been conducted up to this point is that virtually all studies investigating prosody in production have used simple reading or repetition tasks to elicit utterances. The reason is that in order to test prosodic and syntactic hypotheses adequately, it is necessary to precisely control what the speaker says. But unless more naturalistic tasks are used that allow speakers to talk relatively normally, it will be impossible to assess to what extent the need to plan affects the sound features of a sentence, and to evaluate how incrementality in production affects the distribution of hesitations, pauses, and even intonational boundaries across an utterance.

CONCLUSIONS

What is currently known about the process of grammatical encoding indicates that the syntactic structures used in language production have the following characteristics. First, they are generated in two separate stages, the first one creating a representation that represents hierarchical relations but not necessarily linear order, and a second stage in which linearization within phrases takes place. Second, the structures for creating both the global form of the entire utterance and the form of the individual phrases are generated from trees anchored to specific lexical heads. Third, there is some evidence that syntactic representations contain gaps or traces. Admittedly this is a point on which there is little consensus and almost no data, but recent evidence about the computation of subject-verb agreement (Franck et al., in press) as well as data concerning the blocking of function word reduction following a gap suggests that gaps are in fact mentally represented at some stage in production. Moreover, all of these features of grammatical encoding can be captured using TAG as the representational format for syntactic information (F. Ferreira, 2000), which again assumes lexical generation of structure. The main verb of an utterance provides an overall clausal template constrained by the verb's phrase-taking properties, and then each specific phrase is fleshed out and attached as its head (e.g., a noun for a noun phrase) is accessed. TAG represents gaps not via movement but as part of the treelet anchored to the lexical item, thus explaining phenomena such as the blocking of function word reduction in the hypothesized vicinity of a gap.

In addition, although the concept of processing resources is somewhat vague (as MacDonald & Christiansen, 2002 argue), we can resort to an operational definition and say that processing resources are what measures such as initiation time and pause probability/duration reflect. With this assumption, we conclude that grammatical encoding requires resources, and that some constructions appear to be difficult to generate even in felicitous contexts. In addition, the bulk of the evidence indicates that production is incremental in the sense that the most accessible concept will tend to capture the syntactically most prominent position in a functional level structure. This tendency toward moderate incrementality reduces the computational burden on the grammatical encoder because the system can begin with what is already accessible and wait for other elements to become available as processing unfolds. In addition, if it indeed turns out that syntactic priming is particularly robust in dialogue because it makes the task of generating a syntactic structure easier (Pickering & Garrod, 2004; Pickering & Branigan, 1999), then we have further evidence that grammatical encoding requires significant processing resources. Recall that the original argument against this idea was that syntactic processing was assumed to be modular (Levelt, 1989), and one of the characteristics of a module is that it operates automatically (J. A. Fodor, 1983). Do our conclusions undermine this assumption? Not necessarily. Fodor's conception of automaticity appears to have more to do with whether a person's conscious goals and intentions can influence processing than with whether the process is computationally costly. Moreover, it is clear that many specialized systems call upon working memory resources, and one point of debate has been whether the working memory that is involved is domain-general or entirely devoted to just that module. Thus, a system might be modular but still draw on working memory, and the resource pool that is used could itself be modular, in the sense that it is dedicated to processing in that one domain. These are topics for further investigation.

We also conclude that syntactic choices are made largely for the benefit of the speaker. The decision about what syntactic construction to use is at least in part based on the accessibility of the lemmas that will comprise the utterances. Optional function words such as complementizers are left out when the speaker has time to retrieve the immediately following word but included when retrieval is slow and difficult. There is also evidence that speakers obey the Gricean Maxim of Quantity only to a limited extent (Grice, 1975), in part because they have a tendency to describe objects in whatever way is salient to them, neglecting to take into account the effect the description might have on the comprehender (Engelhardt, Bailey, & Ferreira, in press). For example, even though the relevant discourse might include just a single hat, studies show that more than 25% of the time, the hat will be described as the red hat or the hat with the feather. These over-descriptions likely occur because, from the speaker's point of view, the object IS a red hat or a hat with a feather. Because those features of the object are salient, they have a good chance of making it into the message-level representation. In these situations, it would require extra effort for the speaker to produce concise descriptions, because he or she would have to remove content to make sure that information did not get grammatically encoded.

Finally, the syntactic structure of a sentence affects the way it is spoken. For example, the presence of a gap in surface structure affects whether a preceding word is reduced or lengthened. More generally, syntax has profound effects on all aspects of prosody, including the duration and stress level of words, the location and duration of pauses, and the intonational tune and phrasing of the sentence. An unresolved question is whether syntax is directly responsible for these effects, or whether they are mediated through prosodic constituency. Another is how linguistic structure and performance limitations play off of each other to help establish a sentence's overall prosodic form. In addition, it is still not clear how lemma retrieval, word-form activation, and functional and positional level processing are coordinated with the tasks of creating prosodic constituents, generating intonational contours, and implementing a phonetic plan (F. Ferreira, 1993). Moreover, essentially the same questions can be asked about prosody that we considered with respect to syntax in the present review. What sorts of computational resources does the process of creating prosodic representation draw upon, and how do speakers manage and even take advantage of optionality in prosody (for some discussion, see Steedman, 2000; Watson & Gibson, 2004). Unfortunately, although we can ask these questions, there is still not enough evidence to allow us to provide even tentative answers. Ultimately, a complete understanding of syntax in production will require consideration of issues relating to prosody.

We therefore hope that the next decade will see an integration of research on syntax, prosody, and language production.

[This Page Is Intentionally Left Blank]

CHAPTER 3
SPEECH DISORDERS

INTRODUCTION

Speech is a way that use to give a signal, through complex sounds with rapidity and ease, in a variety of different listening situations, and talkers who all speak in their own individual voice (Scott, 2012: 26). In spectrographic displays of the speech signal, the release is indicated by the burst, and the onset of vocal fold vibration by the first vertical striation following an interval of aperiodic energy (Weismer, 2006: 105). According to Fowler and Magnuson (2012: 3), said that speech perception refers to the means by which acoustic and sometimes visual or even haptic speech signals are mapped onto the language forms (words and their component consonants and vowels) that language users know. A speech disorder refers to communication impairment such as stuttering, impaired articulation, and voice impairment, it's referring to difficulties producing speech sounds or problems with voice quality. Weismer (2006: 93) said that speech is an uncontroversial component have been substantial controversies concerning the explanatory role of psycholinguistics in speech disorders that have been concerned with the processes underlying the production, perception, and comprehension of language, in which the speech is an uncontroversial component. According to Weismer (2012: 93), said that psycholinguistic approaches to understanding speech disorders, with a selected focus on developmental speech delay and speech production in persons with neurological disease. Other speech disorders (such as the speech of persons with cleft palate, hearing impairment, and fluency problems, as well-developmental apraxia of speech) have also been conceptualized in psycholinguistic terms, but the two disorders reviewed here provide excellent case studies of the benefits and dangers of the psycholinguistic perspective.

A. Speech disorders include:

1. Art iculation disorders:

Articulation disorders is an impairment in the organization of these phonemes within a language (Bauman-Waengler, 2008). Difficulties producing sounds in syllables or saying words incorrectly to the point that listeners can't understand what's being said.

2. Fluency disorders:

The fluency disorder is referred to speech that is marked by phrase repetitions, interjections, pauses, and revisions like the ones just listed. Fluency disorders can be caused speech output may be non-fluent, hesitations, stops and starts, slow, need more effort to produce, and absence of normal pitch and stress variation. These fluency disorders produced by people who stutter interfere with their ability to communicate effectively and may cause the speakers to have negative emotional reactions to their own speech. These conditions called stuttering, which is the most common form of fluency impairment. The term disfluency is used to describe speech that is marked by phrase repetitions, interjections, pauses, and revisions like the ones just listed. Problems such as stuttering, which the flow of speech is interrupted by abnormal stoppages, repetitions (st-st-stuttering), or prolonging sounds and syllables (ssssstuttering). Stuttering is not the only type of fluency disorder. Some individuals speak not fluently as a result of psychological trauma, neurological disease, or brain injury.

3. Voice disorders:

Problems with the pitch, volume, or quality of the voice that distract listeners from what's being said also cause pain or discomfort for a child when speaking. So, People have a voice disorder, may sound hoarse or breathy *Hypernasality* (Talking out of your nose), *Hyponasality* (Cold and are stuffed up), or have a pitch problem (voice is too high or too low/ too loudly or too softly or may lose your voice when speaking). Voice disorders may be caused by *nodules* (growths like calluses on the vocal cords), *polyps* (swelling like blisters on the vocal cords), the vocal cord paralysis where one or both cords do not move or should open, caused by *paradoxical* vocal fold movement and also, it's maybe caused by *spasmodic dysphonia* that causes the tight sounding voice that sometimes sounds fine (Omori, 2011: 250-252). The voice disorder also can be developed because of allergies, large tonsils or adenoids, smoking, respiratory infections, and poor voice habits.

4. Dysphagia feeding disorders

These include difficulties with drooling, eating, and swallowing.

Speech disorders can get in early childhood or they can be caused by surgery, stroke, an accident, or old age and can have a marked effect upon the ability to communicate in speech or in writing. (Lanier, 2010: 53). According to Lanier (2010: 10) states that there are three categorizations of speech disorders; they are fluency disorders, articulation disorders, and voice disorders. Fluency disorders are related to the smoothness or rhythm of speech. A person with a fluency disorder may hesitate, repeat words, or prolong certain sounds, syllables, words, or phrases. Fluency disorders are especially common among young children. The second type of speech disorder is an articulation disorder. An articulation disorder is achieved through the use of the lips, tongue, teeth, and palate. The third type of speech disorder is a voice disorder. A person with a voice disorder has a problem producing the sounds of speech.

B. Speech Delay

Speech delay is diseases that can be experienced by early childhood which is a condition where children are less able to convey his wishes through talking. Children's ability to speak does not match friends his age so that in everyday activities children experience obstacles. Constraints experienced by children such as children less able to say what he feels or what he wants, the child feels awkward to join in chatting together with friends, and also children tend to be silent. Constraints are also felt by the interlocutor of the child-like parents, teachers, or friends when they want to invite children to talk.

In communication between the two, there is often a misperception, so the opponent talking to children needs to confirm what is meant by the child's words. There is several factors influence children's speech delay, namely relationships family, type of discipline, style of speech, and help from the teacher. Speech development delay could have been caused by a person's physical condition. This kind of a child, who produces errors not typically seen in the course of normal speech sound development, maybe the type who would be diagnosed with developmental apraxia of speech (Gray Weismer, 2006: 94). ‘If there is a speech disorder that has been most influenced by psycholinguistic models, it is certainly the group of developmental disorders referred to as speech delay’ (Robin 1993 in Gray Weismer, 2006: 94).

The person who has difficulty with a small set of sounds typically mastered late in languages that used them to contrastively would clearly be categorized as having speech delay, specifically, one involving called residual errors are usually thought to be the consequence of a delay in speech motor maturation (Weismer, 2006: 94-95). "Psycholinguistic processes mediated between the underlying, phonological forms and their phonetic realizations signaled a subtle shift away from the belief that speech motor maturity played a central role in the sequence of typical speech sound development, and in a speech delay" (Gray Weismer, 2006: 97-98). Children with speech sound production errors: they produce errors because they have a bad representation of the sound category as a result of deficient perceptual and/or categorization skills.

As noted here, efforts to link specific production error types with specific perceptual problems have not been successful (Best, 1994; Kuhl, 1993 in Weismer, 2006:95). Speech delay are thought to the patterns of speech errors in typical development, it makes sense to have a theory that applies broadly to typical and disordered speech sound development above provides compelling evidence for the potential utility of phonological analyses – whether in the form of generative or OT theories in explaining error patterns and selecting a therapy plan for remediating errors in children with speech delay. These theories move past a simple speech motor perspective of speech sound learning and errors, but do not discard the importance of physiological factors in their explanatory apparatus (see Bernstein & Weismer, 2000 in Weismer, 2006: 105).

There are many sources for typical speech sound development, which differ in certain details depending on sampling method, language, and so forth; excellent reviews of this work and references to the extensive older the literature on developmental norms can be found in among others (Grunwell, 1981 &1987, Smit, et.all, 1990, Smit, 1993 and Dodd, 1995 in Weismer, 2006: 95). Also the qualification to this claim that will not be pursued in detail in this chapter, but it should at least be mentioned that efforts to explain developmental speech sound errors on the basis of immature or faulty perceptual mechanisms have a long history in speech-language pathology, but not one with much empirical success (McReynolds, 1988, pp. 422-424 in Weismer, 2006: 95).

The psycholinguistic link is the possibility of immature perceptual mechanisms affecting a child's ability to form proper category features for the contrastive sounds of her or his language. In fact, at least one theoretical strand in the literature on normal speech sound acquisition works from this perspective, so it is easy to see how it could be extended to children with speech sound production errors: they produce errors because they have a bad representation of the sound category as a result of deficient perceptual and/or categorization skills. As noted here, efforts to link specific production error types with specific perceptual problems have not been successful. (Best, 1994; Kuhl, 1993 in Weismer, 2006: 95).

According to dr. Anggia Hapsari, SpKJ a child psychiatrist, in her interview with Kompas.com media (09/11/19), stated if "Speech delay in children is not a diagnosis but a symptom, so in children with speech delay it is an initial symptom of some kind of disorder ". Anggia then explains speech delay divided into two clusters:

1. Functional speech delay disorders: classified as mild, functional disorders that occur due to lack of stimulation or wrong parenting.
2. Non-functional speech delay disorder: is a result due to a receptive language disorder, such as autism or ADHD (Attention Deficit Hyperactivity Disorder) experienced by children.

So, it was important to monitor the child's growth and development consistently. Through stimulation as early as possible by starting to talk two-way interactions between parents and children to help children develop their vocabulary and emotional abilities will also be more developed. Moreover, reduces direct contact with gadgets and televisions.

C. VOT and the Voicing Distinction in speech disorders

VOT is a case that underlying physiology of voicelessness, and its various disruptions in a neurological disease that complicate the mapping of phonetic fact to phonological inference (Weimer, 2012: 118). Speech production research on both normal speakers and speakers with disorders probably no other measure has been explored so consistently as VOT, that typically obtained from the speech acoustic signal, is defined as the time the interval between the release of a stop consonant and the onset of vocal fold vibration for the following vowel. For example, an English-speaking, post-stroke patient who on perceptual analysis produces an apparent speech error of replacing a voiceless stop consonant (such as p, t, or k) with its voiced cognate (b, d, or g) and whose voice-onset time (VOT) for the error is in the short-leg range (say, less than 20 ms) rather than the target-appropriate, long-lag range (greater than 30 ms) might be said to produce a phonemic, not a phonetic error.

On this view, the patient is making an error of selecting the wrong phonological unit, not of misarticulating a correctly selected unit. So it means that the psycholinguistic processes that link phonological units to their phonetic implementation: stop consonants specified as– voice (i.e., phonologically voiceless) are implemented phonetically with long VOTs and stop consonants specified as voice are implemented with short VOTs. When a short VOT is produced for a – voice stop, an inversion of the assumed psycholinguistic process for this phonetic implementation rule leads to the explanation of an incorrectly selected phonological unit. (Weismer, 2012: 93).

When a two- or three-way opposition involves voiced and voiceless unaspirated stops, it is not at all difficult to envision the physiology of voiced stops resulting in a short-lag VOT and thus the possible interpretation of a phonemic (voicless for voiced) error. With no LDG but vocal folds that are prevented from vibrating because the pressures below and above the glottis are the same, as often occurs toward the end of voiced stop closure interval, the vocal fold vibration may be delayed following the stop release by as much as 20 ms, resulting in the kind of VOT observed for voiceless unaspirated stops. English in using other parts of the VOT continuum to signal the contrast: voiceless stops are unaspirated and produced with short-lag VOTs, voiced stops with voicing lead. No matter where the contrast is made along its continuum, VOT corresponds consistently well with the phonological voicing distinction in many different languages.

In fact, because of Lisker and Abramson’s work and its many *sequela* VOT is often regarded as a good expression of the voicing status of stop consonants it's this correspondence the assumption of a fairly straightforward mapping between the acoustic measure and phonological voicing status that prompted literature on voicing errors in neurogenic speech disorders and inferences from acoustic data concerning their origin (Weismer, 2006: 106). "The actual separation of VOT values for different members of a phonological voicing contrast is somewhat messier than implied by this brief description, as a result of factors such as phonetic context, position in the word, stress, speaking rate, speaking style, and even speaker age. The discussion here is limited to those cases in which the VOT-voicing status mapping is thought to be clear, for example, stops produced in word-initial, stressed syllables spoken at conversational rates and in non-casual speaking styles" (Weismer, 2006: 106).

"The VOT measurements of the posterior aphasics revealed only the presence of phonemic substitutions, i.e., clear-cut shifts in phonetic category from the target phoneme these errors reflect a deficit in selecting the appropriate phoneme or underlying phonological form, and subsequently programming correctly the articulatory commands for the substituted phoneme" (Blumstein et al., 1977: 381 in Weismer 2006: 107). VOT in the short-lag region is shown to be a possibility, and that possibility has been interpreted at times as a voiced for the voiceless error of the phonological kind. These hypothetical cases, however a physiologically plausible and are clearly phonetic anomalies. For these reasons, when a voiceless stop target is produced with a short-lag VOT it seems ill-advised to interpret the event as evidence of a phonological error, at least in the absence of additional evidence.

The VOT interval is clearly much shorter for /d/, as compared to /t/, and these typical short-lag VOTs for voiced stops will occur whether or not voicing continues throughout the closure interval or is terminated before the end of the closure interval. The thing that important here, and critical to the issue of inferring phonological status from empirically measured VOT values, is that this underlying physiology can be boiled down to a fairly clear dichotomy: voiceless stops have an LDG, voiced stops do not. When the LDG is present, at least for intervocalic, prestressed stops, VOTs will be in the long-lag range simply because the supralaryngeal closure interval is released about halfway into the LDG when the vocal folds are maximally open (Löfqvist, 1980 in Weismer, 2006: 111).

Conversely, when there is no LDG, as is the case for the upper right panel of Figure 2, periodic motions of the vocal folds can resume shortly after the release of the closure interval. Thus, in an English speaker with no neurological disease, the presence of the LDG can be assumed to be part of the voiceless specification for stops, and its absence part of its voiced specification. *Presumably*, a correct phonological representation of a /t/ will trigger the LDG as part of its a specification for articulatory control; a /d/ will not. Note the bottom panels of Figure 2, where the typical physiology is illustrated for /t/ (bottom left) and /d/ (bottom right) in utterance-initial position, as would be the case for isolated words with initial stops.

Even though there is no voiced segment preceding the /t/ an LDG is observed. For utterance-initial /d/, especially in English, there are typically no vocal fold vibrations during the closure interval (i.e., utterance-initial voiced stops are typically devoiced in English), but the vocal folds held in the midline position during the closure and so begin to vibrate shortly following the stop release. The binary physiological opposition of presence vs. absence of the LDG applies equally to utterance-initial stops, as it does to intervocalic stops, as important point as some experiments use isolated words to study the voicing opposition in aphasia or other neurogenic speech disorders (Weismer, 2006: 111-112). Like most generalizations of this sort, there are some qualifications; as vocal fold motions approach the closure interval of a stop consonant, details of the vibration are somewhat different for voiceless vs. voiced stops (Ní Chasaide and Gobl, 1997 in Weismer, 2006: 111).

The best form of such evidence would be direct viewing of the larynx during the production of voiceless (Weismer, 2006: 113). VOT values in the ‘lead’ (negative) range for the voiced function. In connected speech, when voiced stops are in the intervocalic position vocal fold vibration can continue throughout the closure interval, with the duration of voicing lead equal to the duration of the closure interval. For this analysis, if voicing occurred at the same time as the stop release (which will be the case, give or take several milliseconds), the VOT was recorded as ‘0’ (zero) (Weismer, 2006: 116).

D. Apraxia

Apraxia is defined as the difficulty or inability to perform learned skilled actions. Identifying apraxia in patients has prognostic implications (Rajan, 2018: 1). Apraxia in behavioural neurology refers to the loss of the ability to carry out learned, skilled actions in the absence of motor, sensory, coordination, or comprehension abnormalities (Rothi and Heilman, 2003 in Rajan, 2018: 1). It is to be differentiated from akinesia, which is defined as a general failure to initiate movement in the absence of weakness (Heilman and Watson, 2008 in Rajan, 2018: 1). Apraxia is a helpful localising sign on the mental status examination and often predicts disability in patients with stroke or dementia. It can affect both sides of the body, even when the underlying lesion is unilateral (Rajan, 2018: 1). “Individual testing of components of apraxia such as conceptual, conduction, visuoimitative and dissociation apraxia is beyond the scope of this review; below, we aim to provide ‘practice pearls’ for the general clinician and interested student.

Reliable testing of apraxia depends on the patient’s ability to understand commands and move the limbs without weakness” (Rajan, 2018: 8). “Apraxia, particularly when present with other deficits, can help localise lesions, and hence it is useful to test praxis of individual limbs, the buccofacial region and axial structures separately. Transitive and intransitive gestures are tested separately. As discussed earlier, apraxia often coexists with aphasia, which may impair a patient’s ability to understand commands. Severe limb apraxia may be associated with impairment in gestures, and severe orofacial apraxia may be associated with impaired verbal communication. Agnosia and spatial neglect are also often associated with apraxia, especially in strokes, and this may significantly impair accurate assessment of apraxia as well” (” (Rajan, 2018: 8). Apraxia of speech in the acute setting of stroke is commonly misdiagnosed as aphasia. Detailed testing at the bedside can be difficult, but if the patient’s writing and reading/auditory comprehension are normal, and speech is notable for phoneme prolongation and inter-syllabic segmentation, then apraxia of speech rather than aphasia should be considered (Polanowska and Pietrzyk-Krawczyk, 2016 in Rajan, 2018: 8).

People who have Apraxia often recognize their errors when speaking and try to correct them resulting in "groping". Because, actually the errors of Apraxia are variable, sometimes people with Apraxia may say a word correctly and other times not, also have more errors on consonants than vowels. they know what they want to say, but can't say it and also maybe they able to write it down or give descriptions. According to Lanier (2010: 30-31), said that apraxia is caused by speech disorder caused by motor brain damage and can affect articulation, fluency, voice disorder (or the combination). Blumstein et al. (1977, 1980) in Weimer (1997) said if the vocal folds are close to, or at the midline when a stop is released, it does not seem possible for VOT to be much more than 20, or at the limit, 30 ms; a voiced target produced with a VOT of 50–60 ms would seem to require an LDG which, as argued above, signifies voicelessness. Perhaps, then, measurement of a long-lag VOT when the target stop is voiced should be taken as good evidence of a phonemic error. There is, however, one more piece of evidence that demonstrates how tenuous this interpretation might be.

There are two types of apraxia that are developmental apraxia of speech (DAS) and acquired apraxia. DAS usually occurs in children since they are born, the characteristics of babies who have it are: they become quiet babies, they don't do babble as infants, and the first words are usually delayed compared to the normal time and when they grow up, they do not have problems with listening and understanding others' speech. Besides that, acquired apraxia is usually found in adults and its cause by the impairment of their ability to speak. This inability results from illnesses like a stroke, head injury, or tumor. The characteristic is that people who suffer from it have difficulty in constructing their speech in the correct arrangement when expressing what they think. Therapy for apraxia is still experimental, and much of the evidence for targeted rehabilitation comes from studies on stroke patients. Compared to conventional rehabilitation for aphasia, a behavioral training program of gestural exercises has been shown to improve limb apraxia specifically and functional independence generally (Smania et al., 2006 in Rajan, 2018: 7).

E. Dysarthria

The Dysarthria is a group of impairments that may affect the speed, range, direction, strength, and timing of motor movements. That the errors are consistent, distortions, and omission of sounds. Dysarthria is a neurogenic speech disorder in which damage to the central or peripheral nervous system results in a problem with control of some or many of the scores of muscles involved in the production of speech (Weismer, 1997). According to Lanier (2010: 27), dysarthria is a group of neurologically related speech disorders. Known as motor speech disorders, caused by lesions on the brain in areas responsible for planning, executing, and controlling the movements necessary for speech. The affected speech muscles become weak or paralyzed.

Dysarthria is most common in people born with cerebral palsy (CP) or muscular dystrophy and adults who have experienced a stroke, tumor, or degenerative such as Parkinson's disease. The diseases that result in dysarthria may also produce cognitive problems (such as mental retardation in cerebral palsy, dementia and depression in Parkinson disease, aphasia in stroke, to name a few examples), the speech motor control problem has never been thought to be complicated by a potential loss or modification of phonological representations. Stated otherwise, sound segment errors and their acoustic manifestations in dysarthria have always been considered of a phonetic origin, reflecting only the control problem (Weismer, 1997).

Thus, studies of speech production in aphasia can be added to those in dysarthria, for example, to form a global view of how neurological damage affects speech motor control. In this regard, it would be very useful to obtain more relevant data in a variety of languages with varying phonological and phonetic characteristics (Weismer, 2006: 118). Phonetic anomalies also occur in dysarthria, a group of disorders in which phonological representations are assumed to be unaffected by the disease process, highlights another problem in the use of VOT values to make the distinction between phonetic and phonological errors (Weismer,2006: 117).

Speech samples produced by 22 adults with dysarthria (Weismer, 2004); none of these speakers had serious cognitive problems, and all had been seen for research purposes related solely to their speech motor control deficit. This cumulative probability graph shows VOT for voiced (solid function) and voiceless (dotted function) stop consonants produced in the prestressed position. For 97 voiced stops and 112 voiceless stops, median VOTs across speakers and place of articulation were 16 and 49 ms, respectively, with both functions’ steeper below, as compared to above, the medians.

From the above explanation it can be concluded if a speech disorder refers to communication impairment such as stuttering, impaired articulation, and voice impairment, it's referring to difficulties producing speech sounds or problems with voice quality. Weismer (2006: 93) said that speech is an uncontroversial component have been substantial controversies concerning the explanatory role of psycholinguistics in speech disorders that have been concerned with the processes underlying the production, perception, and comprehension of language, in which the speech is an uncontroversial component. So, speech disorder affects the way a person talks. Speech disorders can affect a person's self-esteem and overall quality of life. There are many types of speech disorders include stuttering, apraxia, and dysarthria.

The possible causes of speech disorders can be muscle weakness, brain injuries, degenerative diseases, autism, and hearing loss. People who have speech disorders usually know precisely what they want to say and what is appropriate for the situation but there has trouble in producing certain sounds to communicate it effectively and accurately, it's the same as the production of vocal sounds without word-formation.

CHAPTER 4
LEARNING THE SOUNDS OF LANGUAGE

INTRODUCTION

The infant's world is filled with sound. How do infants begin to make sense of their auditory environment? To do so, infants must discern what is most important in the rich and complex acoustic signal that they receive as input. Infants, however, seem to master this task with remarkable ease. By the end of the first year of postnatal life, infants are sophisticated native language listeners, with knowledge about how individual sounds work in their language and how those sounds combine to make words. Moreover, this ability to process complex linguistic input is not limited to the auditory modality; infants exposed to signed languages exhibit a similar trajectory of language discovery. Even infants exposed to multiple languages somehow appear to discover the structures that are most relevant in each language. How do infants converge on their native language system(s) so rapidly? Historically, the answer to this question was that infants, like adults, treat speech as "special" – that is, as a privileged auditory input distinct from all other perceptual stimuli (e.g., Eimas, et al., 1971). In this view, our brains use specialized processes designed to operate over speech stimuli. This account has spurred decades of research focused on the degree to which the processes subserving speech perception and language acquisition are specific to speech per se, or available more generally for perceptual processing across multiple domains and species (for a current incarnation of this debate, see the exchange between Fitch, Hauser, and Chomsky, 2005 and Pinker and Jackendoff, 2005). What is clear is that the spoken environment (and the signed environment, for infants exposed to signed languages) presents a serious challenge for infant learners. The linguistic structures to be acquired are massively complex, containing multi-tiered layers of information and numerous probabilistic cues. To figure out what matters in the input, infants presumably rely on some set of innate structures (a.k.a. "nature"), which may include learning mechanisms (complete with input representations and computational algorithms) and perceptual biases (inherent preferences for some stimuli over others). This preexisting machinery is then coupled with a rich environment (a.k.a. "nurture"), which includes linguistic, referential, social, and affective information. In this view, learning is an essential bridge between nature and nurture: The learning process makes use of structures internal to the infant to organize incoming information, thereby shaping how new input is processed and represented. In this chapter, we will briefly review recent developments in the study of how infants learn to perceive the sounds of their native language, focusing on how infants learn about individual native language sounds, how those sounds are combined, and the beginnings of word-level knowledge. We will emphasize what is currently known not just about what babies know and when they know it, but how that knowledge is acquired.

A. Different Languages and Different Sounds

It is often said that languages differ by sound or melody. What does this mean? Only when one begins consciously the process of learning a foreign language, does one notice that the language in question possesses sounds far removed from those in one's own, and not even produced in the same manner. Sometimes, there are also sounds which sound similar, yet prove to be different by a minute, but essential, detail. Those sounds cannot simply be replaced by sounds one knows from their own language. Such a replacement could change the meaning of a word or phrase, or even cause the sentence to become incomprehensible. Correct articulation can prove to be of great difficulty and may require arduous and repetitive practice. Several different sounds may sound the same to a non-native speaker, and at the same time, deceptively similar to a sound from their own mother tongue. Although awareness of such phenomena increases with every new foreign language learnt, only a few realise just how much variety of sound exists in the languages of the world.

B. Strange Sounds of Strange Languages

'Strange' is of course a term used half-jokingly here and should be understood rather as 'rare'. The readers of this text are probably familiar only with major European languages, thus they might find some of the articulation phenomena or sounds existing in languages of Asia and Africa quite different from their own. It is often assumed that vowels are always voiced as that is the case in the most commonly spoken languages in Europe. When whispering, however, the vowels of the aforementioned languages can be rendered voiceless, all the while remaining comprehensible and distinguishable.

Nevertheless, such cases occur rarely. Languages whose sound systems comprise voiceless vowels include American Indian languages like Zuni or Cheyenne, or Japanese. Laryngealization can be regarded as a particular voice quality. It is realised as a kind of croak, especially at the end of sentences. It appears when the vocal folds vibrate irregularly as a result of low volume-velocity of the air flow. There are, however, languages such as Kedang (an Austronesian language spoken in Indonesia) or Jalapa Mazatec (spoken in Mexico), where a creaky voice is an important sound quality. Although aspirated consonants are quite popular and exist in many languages of the world, aspirated vowels present in Jalapa Mazatec or Gujarati, remain quite a rarity.

C. How to Transcribe Sounds of A Language?

There is often no correlation between the graphic signs of a language and its sounds in many languages of the world. Their written symbols frequently do not pertain to the actual sounds of the written texts. They may represent whole words and thus, they do not show the reader of what sound units those words comprise. One could expect the highest level of correspondence between graphic signs and sounds from languages based on an alphabetic script, such as Latin (which, in one form or another, is used by most European languages). Even here, however, one encounters difficulties as one letter may signify different sounds and be read in various ways. In English, a double *o* is pronounced differently in *blood*, *book* and *door*. The fact that sometimes up to four letters are used to represent a single sound (for example, in *th**ough***) may also leave one wondering. Writing systems of natural languages are usually based on tradition and in many cases they prove difficult for linguists to analyse. Moreover, many languages, especially the less widely used or endangered ones, do not even have a writing system. Therefore, there was a need for a universal, well-planned system for writing down the sounds of speech. This system could be used equally well to write utterances in a known as well as an unknown language, as it would be capable of noting down sounds of almost any thinkable manner of articulation. The system itself is quite complex and its use requires a lot of time and effort spent on mastering it. Even experienced phoneticians may not always agree as to the phonetic transcription of a particular utterance, because the system operates on binary and absolute categories. A given sound can either be transcribed as voiced or voiceless, nasal or non-nasal. In practice, however, a phonetician knows that nasality or voicing are gradable in actual speech and it might be difficult to decide whether the feature has already appeared or not (i.e. if the given segmental is voiced or not).

Knowing the phonemic inventory of a given language (i.e. the set of its 'basic sounds'), it is possible to transcribe speech in a less complicated manner. A broad (or phonemic, phonological) transcription takes into consideration only the phoneme which the particular sound of the utterance belongs to. In this case, the articulatory details of the speaker are irrelevant. Only the key features matter that differentiate the categories of sound. The transcription only contains the number of symbols needed to transcribe all phonemes of a given language. It is, thus, language-related, although at times it can serve a larger number of languages with similar phonological systems. A native speaker can use it without the arduous training required to master narrow phonetic transcription. It can still prove difficult to a person who does not know the language in question as she may not be familiar with what the important features differentiating segments in the given language are. Neither will she know how to assign them to particular phonemes.

In view of some of the unusual symbols of the IPA, some phoneticians decide to use different symbols, based on combinations of letters known from the Latin alphabet. This system is known as SAMPA. It is most often used for broad transcription. SAMPA has been adapted to many languages and thus, there are various 'national' variants of this kind of notation. When necessary, however, linguists may use transcription systems based on the orthography of the given language. It can be extended with additional symbols in order to transcribe phenomena such as yawning, silence, interjections or to specify the voice quality (i.e. creaky, high, whisper). Some decide not to include punctuation, capital/small letter distinction (if such a distinction exists in the language) or other orthographic rules so as to minimise arbitrariness and subjectivity. Transcription, however, always remains a subjective interpretation of speech.

Here is an example using three different manners of transcription – orthographic, IPA and SAMPA for a phrase "Pewnego razu Północny wiatr i Słońce sprzeczali się". To facilitate comparison, the short text was divided into syllables. It is, however, a broad, hypothetical transcription as it can be realised differently (for example by keeping the last segment nasal).

1. Background issues

a. Early research in infant speech perception

The field of infant speech perception took off in 1971 with the publication of a seminal paper by Peter Eimas and his colleagues (Eimas et al., 1971). Prior to this study, little research had applied the constructs of adult speech perception to investigate the development of speech perception – no methods were available to study the internal perceptual representations of prelinguistic infants. In the first major study of its kind, Eimas and his colleagues were able to demonstrate a remarkable ability on the part of very young infants (one- and four-montholds). The syllables /pa/ and /ba/ vary along a dimension known as voice onset time (VOT): the time between the release of the consonant and the opening and closing of the vocal folds. This timing varies reliably across consonants; the VOT for /p/ is longer than the VOT for /b/. Earlier studies with adult participants had demonstrated a phenomenon known as categorical perception, which appeared to be specific to speech (e.g., Liberman et al., 1967). Adults reliably discriminated differences in VOT only when the sounds in question spanned two categories (such as /b/ versus /p/). Using a nonnutritive sucking methodology, in which the rate at which infants sucked on a pacifier determined the presentation of auditory stimuli, Eimas et al. (1971) demonstrated that infants showed equivalent discrimination abilities: good discrimination of token pairs that spanned a category boundary, but poor discrimination of token pairs within a category. These results suggested that, like adults, infants perceived phonemes categorically. The fact that even such young infants demonstrated this hallmark of speech perception is consistent with the hypothesis that the "specialness of speech" previously observed in studies with adult participants might be present from birth. The second major development in the study of infant speech perception arose from the comparison of younger and older infants. Younger infants, like those studied by Eimas et al. (1971), appeared able to apply their special speech perception abilities, including categorical discriminations, to sound contrasts from many languages of the world, including sound contrasts that did not occur in their native language (e.g., Lasky, Syrdal-Lasky, and Klein, 1975; Trehub, 1976). By contrast, groundbreaking research by Werker and Tees (1984) demonstrated that older infants (ten to twelve months of age) were far less open-minded in the application of their speech perception abilities. By the end of their first year, infants were able to discriminate only those consonant contrasts present in their native language, with a few notable exceptions, such as Zulu clicks (e.g., Best, McRoberts, and Sithole, 1988). The same pattern of initial perception of nonnative contrasts, followed by a tuning of the speech perception system, was observed at a somewhat earlier age for vowel contrasts, by six months of age (Kuhl et al., 1992). This developmental difference between vowels and consonants presumably reflects the fact that vowels are particularly clear and emphasized in infant directed speech (Kuhl et al., 1997). These results suggest the presence of a powerful mechanism underlying reorganization of perceptual systems as a function of linguistic experience (for recent reviews, see Kuhl, 2004; Werker and Curtin, 2005). Importantly, this tuning of the perceptual system to fit the native language occurs before infants produce their first words – around twelve months of age on average, with great variability (e.g., Fenson et al., 1994). Recent evidence suggests that similar reorganization as a function of perceptual learning occurs during infancy in another domain of auditory experience: musical rhythm (Hannon and Trehub, 2005).

As in the linguistic domain, this musical "tuning" occurs well before the onset of infants' productive use of the system, suggesting a process of perceptual learning that is at least somewhat independent of production.1 Further evidence to suggest that infants are extremely skilled at extracting structure from speech comes from the domain of word segmentation. Speech is typically fluent, and does not contain reliable breaks between words analogous to the white spaces in text; even speech to infants does not contain pauses or other acoustic markers of word boundaries. This raises a potential problem for infants, who must somehow break into the speech in their linguistic environments to find word boundaries. In a seminal paper, Jusczyk and Aslin (1995) demonstrated that after hearing target words in passages (e.g., "The cup is filled with milk. Mommy's cup is on the table."), seven-month-old infants subsequently listened longer to the familiar target words (e.g., cup) than to words that were not contained in the familiarization passages (e.g., bike). These findings provided the initial experimental evidence demonstrating that infants can segment word-like sound sequences from fluent speech. Notably, this task is sufficiently difficult that computer speech recognition systems can only segment words from fluent speech following explicit training and instruction. Nevertheless, infants appear to perform this task with ease, and certainly do not require explicit training or feedback. As with the development of phonemic perception, this process clearly occurs independent of word production, and also largely independent of knowledge of the meanings of the newly segmented words – though it is certainly the case that segmentation feeds upon itself; familiar words provide powerful word boundary cues for adjacent novel words (e.g., Bortfeld et al., 2005). Thus, by the mid-1990s, it was clear that infants can perform complex operations over the speech they hear in the environment. Learning, coupled with potential innate predispositions, was evidently a powerful driving force at the outset of language acquisition. In the remainder of this chapter, we will consider some of the important developments in this area, focusing particularly on the mechanisms that subserve infants' early attainments in becoming native language listeners.

b. Development of speech perception

From the earliest moments of postnatal life, speech– and the human faces that produce it– is a central feature of infants' environments. Indeed, infants begin to discover regularities in speech prior to birth, via sound transmission through the uterine wall. Prenatal exposure provides sufficient information to permit newborn infants to distinguish the prosody (pitch and rhythmic structure) of their mother's language from other languages (Mehler et al., 1988). These musical properties of language appear to be particularly interesting and useful to very young infant perceivers. For example, newborn infants can distinguish between languages that have different characteristic rhythmic patterns (e.g., English versus French), but not those that share rhythmic patterns, such as English versus Dutch (Nazzi, Bertoncini, and Mehler, 1998). Interestingly, both infants and nonhuman primates show evidence of this focus on rhythm in their pattern of discriminating between rhythmically dissimilar languages in forward speech, but not backward speech, which disrupts rhythmic cues (Ramus et al., 2000; Tincoff et al., 2005). Only by five months of age does experience with their native language allow infants to discriminate it from other rhythmically similar languages (e.g., Nazzi, Jusczyk, and Johnson, 2000). This developmental trajectory suggests that we begin postnatal life attending to rhythm in quite a general way, akin to other primates, and that with experience, we learn to attend to rhythm as it functions in our native language. Speech clearly has great perceptual significance to infants, and is certainly special in that sense. For example, newborns, two month-olds, and four-month-olds prefer to listen to speech as compared to nonspeech analogues that share many critical acoustic features with speech (Vouloumanos and Werker, 2002; 2004).

The brains of young infants also treat speech and nonspeech differently, with greater left hemisphere activation for speech (Dehaene-Lambertz, Dehaene, and Hertz-Pannier, 2002; Peña et al., 2003). Moreover, infants are particularly interested in the kind of speech that adults typically direct to them (e.g., Cooper and Aslin, 1990). Infant-directed speech (IDS) has characteristic features that emphasize its musical components: IDS is higher pitched, slower, and contains larger pitch excursions than adult-directed speech, as measured across a number of different language communities (e.g., Fernald, 1992; Gleitman, Newport, and Gleitman, 1984; Liu, Kuhl, and Tsao, 2003). In addition, the individual sounds in IDS are exaggerated relative to adult-directed speech (Kuhl et al., 1997), and the clarity of maternal speech is correlated with infant speech perception abilities (Liu et al., 2003). Infants' auditory environments and infants' perceptual systems appear to be well-suited to one another, facilitating the learning process. While speech itself may be a privileged stimulus for infants, the operations performed over speech during language acquisition do not appear to be specialized for speech. For example, consider categorical perception. Once considered a hallmark example of the specialness of speech, categorical perception has since been demonstrated for other stimuli that share the temporal dynamics of speech but that are not perceived as speech, including tone pairs that varied in their relative onset, mimicking the acoustic timing characteristics of VOT (for infant data, see Jusczyk et al., 1980; for adult data, see Pisoni, 1977).

In addition, categorical perception has now been demonstrated across many other domains, from musical intervals (e.g., Smith et al., 1994) to facial emotion displays (e.g., Pollak and Kistler, 2002; for review, see Harnad, 1987). As in these nonlinguistic domains, listeners are also sensitive to some within-category differences in speech sounds both during adulthood (e.g., McMurray, Tanenhaus, and Aslin, 2002) and infancy (McMurray and Aslin, 2005). Additional evidence that many of the operations performed over speech are not specialized for speech processing comes from studies using nonhuman participants, including primates, small mammals, and birds. While the auditory systems of nonhuman animals share many features with our own, these other species presumably did not evolve perceptual systems adapted to process human language. Despite this fact, nonhuman mammals (Kuhl and Miller, 1975, 1978; Kuhl and Padden, 1983) show categorical perception for both voicing (e.g., /pa/– /ba/) and place (e.g., /ba/–/da/) continua, and birds show a warping of their perceptual space to reveal a prototype organization following exposures to vowel categories (e.g., Kluender et al., 1998) as previously seen in human infants (e.g., Kuhl et al., 1992). The body of data from nonhuman animals supports the hypothesis that many aspects of speech perception may have developed to take advantage of existing perceptual discontinuities and learning mechanisms not unique to humans (e.g., Kluender et al., 2006). Regardless of one's theory of the origins of speech processing in infancy, probably the most remarkable thing about infant speech perception is the pace of the learning that occurs: By six months of age for vowels, and ten months of age for consonants, infants have learned to categorize speech sounds as demanded by the categories present in their native language. How does this learning process unfold? Given the precocity of the nonnative to native language shift in speech processing, it cannot be due to factors like the acquisition of minimal pair vocabulary items (e.g., using the fact that /pat/ and /bat/ have different meanings as evidence that /p/ and /b/ are different phonemes).

For this to be the case, we would expect to see extensive vocabulary learning preceding the loss of native language speech sounds. Infants do comprehend some words during the first year (e.g., Tincoff and Jusczyk, 1999). However, it is highly unlikely that infants learn enough words, and, critically, the right kinds of words (such as minimal pairs), to drive their early achievements in speech perception. One potentially informative source of information pointing to speech categories lies in the distributions of individual sound tokens. Recent analyses and modeling support the intuition that infant-directed speech contains clumps of exemplars corresponding to native language categories (Vallabha et al., 2007; Werker et al., 2007). Moreover, infants can capitalize on these regularities in lab learning tasks employing an artificial language learning methodology. Maye, Werker, and Gerken (2002) demonstrated that infants' speech categories can be modified by the distribution of speech sound exemplars. When infants heard a bimodal distribution (two peaks) of speech sounds on a continuum, they interpreted this distribution to imply the existence of two speech categories. Infants exposed to a unimodal distribution (one peak), on the other hand, interpreted the distribution as representing only a single category. These results are remarkable both because they show that young infants possess sophisticated statistical learning abilities – namely, the ability to track and interpret frequency histograms as unimodal or bimodal – and because they suggest that infants are able to represent detailed differences between tokens of the same phoneme in order to generate these histograms. This learning is predicated upon the ability to detect extremely small differences between sounds, a point often overlooked given the characterization of speech perception as categorical (for related evidence, see McMurray and Aslin, 2005). To the extent that infants rely on a statistical learning mechanism of this kind, we have an explanation for how infants adjust their phonemic categories well prior to word learning; according to Maye et al.'s (2002) account, top-down knowledge about words is not needed to drive the reorganization of speech perception in infancy. Recent research suggests that basic Hebbian learning mechanisms could be responsible for maintenance of these established perceptual categories, and for shifting from single-category representations to multiple-category representations (McCandliss et al., 2002).

Infants' early speech perception skills are linked in important ways to later language learning. For example, measures of speech perception in infancy predict later native language outcomes, such as vocabulary size, when infants are assessed months or years later (Kuhl et al., 2005; Tsao, Liu, and Kuhl, 2004). These fine-grained perceptual abilities also appear to be maintained when infants are processing familiar words, allowing infants to distinguish correct pronunciations of known words, such as baby, from mispronunciations like vaby (e.g., Swingley and Aslin, 2000; 2002). However, the process of learning new words may obscure infants' access to phonetic detail, at least under some circumstances (e.g., Stager and Werker, 1997; Werker et al., 2002). When infants' new lexical representations include phonetic detail, when they do not, and the factors that influence this process are a matter of current active research, with important implications for how speech perception relates to subsequent language acquisition (for recent reviews, see Saffran and Graf Estes, 2006; Werker and Curtin, 2005). Interestingly, there are features of the interpersonal situation in which infants learn language – which are typically overlooked in experimental studies of infant speech perception – that also appear to matter with respect to the development of speech perception.

In a recent study, Kuhl, Tsao, and Liu (2003) demonstrated that social interaction facilitates learning. To do so, they exposed English-speaking infants to speakers of Mandarin, and then tested them on Mandarin consonant contrasts not present in English. Critically, these infants were of an age (nine months) at which perception of nonnative speech contrasts has typically declined. As seen in some nonhuman species (such as zebra finches), only those infants exposed to the speech stimuli in the context of dynamic social interactions learned the contrast; infants who heard the same speech presented via high-quality DVD recordings failed to learn the contrast, despite being able to see and hear the same input as the infants in the face-to-face interactive context. These findings suggest that the presence of an interacting human may facilitate learning, potentially via attentional modulation, affective engagement, or some other feature of the social environment. Goldstein, King, and West (2003) suggest that at least for the development of infant babbling, interaction with adults may provide an important source of reinforcement in shaping infants' subsequent productions. Future research will need to address how social cues are integrated with the kinds of statistical information shown to be relevant in learning, since infants are able to profit from the structure of the input in some noninteractive contexts (see Kuhl, 2004 for discussion).

c. Beginnings of language: Perception of sound combinations

During the first year of life, as infants are busy figuring out which sounds are meaningful in their language, they are also engaged in discovering the patterns that occur over multiple sound units. These patterns occur at many levels (i.e., patterns of syllables cohere into words, patterns of words cohere into grammatical units, etc.). Many of the patterns that are the initial focus of infants' attention are patterns of sounds that are visible to the naked ear – that is, patterns that have salient acoustic structure. From this starting point, infants can continue on to discover more abstract patterns. One pattern found in many languages that is accessible to infant listeners is an alternation between stressed (strong) and unstressed (weak) syllables, which creates an audible rhythm in many languages, including English. These stress differences are carried by the pitch, amplitude (loudness), duration, and vowel quality (in some languages) of the syllables. Bisyllabic words in English are predominantly trochaic – that is, they consist of strong syllables followed by weak syllables, as in the words doggie, baby, and mommy (e.g., Cutler and

Carter, 1987). If one is familiar with this pattern of correlations between a syllable's position within a word and the likelihood that a syllable will receive stress, these lexical stress cues can be used to segment words from fluent speech. By nine months of age, infants have already learned something about the predominant stress pattern of their native language (Jusczyk, Cutler, and Redanz, 1993; Polka, Sundara, and Blue, 2002). Moreover, infants can use this information: When given a stream of fluent speech made up of nonsense words, nine-month-old infants can use stress patterns alone to segment words (Curtin, Mintz, and Christiansen, 2005; Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003), and can integrate stress patterns with other distributional information (e.g., Morgan and Saffran, 1995). Further, infants are able to learn new stress patterns quickly. After just a few minutes of exposure to a novel stress pattern, such as iambic (weak–strong) stress, six-and-a-half- and nine-month-old infants generate iambic word boundary expectations when segmenting words from fluent speech (Thiessen and Saffran, 2007). Infants are also sensitive to regularities that occur in specific sound combinations. Phonotactic patterns define which sound sequences are legal or likely in a given language.

For example, English words can end with /fs/, but cannot begin with /fs/. These patterns vary cross-linguistically, and are thus not solely determined by such factors as articulatory constraints and pronounceability. By nine months of age, infants have learned enough about the phonotactic regularities of their native language to discriminate legal from illegal sequences (Jusczyk et al., 1993). This knowledge is sufficiently specific to allow infants to discriminate legal sequences that are frequent from those that are infrequent (Jusczyk, Luce, and CharlesLuce, 1994). Knowledge about the likelihood of phonotactic sequences that occur at word boundaries versus word-medially also plays a role in helping infants to discover word boundaries in fluent speech (Mattys and Jusczyk, 2001; Mattys et al., 1999). Phonotactic patterns characterize possible words in a language – sequences with good English phonotactics are possible words, whether or not they have been "chosen" as actual, meaningful words by a language community. Other types of patterns, which are a subset of phonotactic patterns, characterize the actual words that exist in a language. For example, while the sequence pibay is a possible English word, it is currently not attested, and reflects a gap in the English lexicon. Infants are able to exploit the differences between particular patterns of phonemes that are likely to occur in a language and patterns that are not when attempting to segment words from fluent speech. To do so, infants make use of sequential statistics in the speech stream. Saffran, Aslin, and Newport (1996) asked whether eightmonth-old infants could take advantage of these regularities in a word segmentation task. In this study, syllable pairs contained either high or low transitional probabilities. A high transitional probability indicates that one syllable is highly predictive of the next (e.g., the probability that pi will follow ha, as in the word happy).

A low transitional probability indicates that a number of different syllables can follow the given syllable (as in pibay, which does occasionally occur in fluent speech, such as when the word happy is followed by the word baby, but which is unlikely in English). These probability differences were the only cue to word boundaries in the nonsense language used by Saffran et al. (1996). Despite the paucity of input and the difficulty of this task, infants were able to use these sequential statistics to discover word-like units in fluent speech (see also Aslin, Saffran, and Newport, 1998). Similarly, training studies have shown that infants can acquire other types of regularities from brief exposures. Along with novel stress patterns (Thiessen and Saffran, 2007), infants are able to rapidly learn new phonotactic regularities in lab learning tasks (Chambers, Onishi, and Fisher, 2003). They can then use these newly learned phonotactic regularities as cues to word boundaries in fluent speech (Saffran and Thiessen, 2003). Training studies like these demonstrate that infants' strategies based on native language regularities are flexible, allowing learners to take advantage of new patterns in the speech stream following very brief exposures. These findings, along with many others, suggest that language learning is a highly dynamic process, with subsequent learning shaped by prior learning (for a recent review, see Gómez, 2006). The dynamic processes children use to learn language are not limited to discovering patterns. For example, infants take advantage of previous knowledge when trying to learn language. Once six-month-old infants have segmented a word (such as their own name), that word itself becomes a segmentation cue; infants can now segment words that are presented adjacent to their names in fluent speech (Bortfeld et al., 2005). Similarly, attention-getting features of language assist learners. For example, infants in segmentation tasks are able to take advantage of the infant-directed speech that adults use when speaking to them. In a recent study, Thiessen, Hill, and

Saffran (2005) showed that seven-month-old infants found it easier to segment words from infant-directed nonsense speech than adult-directed nonsense speech. Infants heard sentences of nonsense speech that contained pitch contours characteristic of either infant-directed or adultdirected speech. In both conditions, the only cue to word boundaries was the transitional probabilities between syllables, which were uncorrelated with the pitch contours. Participants in the infant-directed condition showed evidence of word segmentation while those in the adult-directed condition did not. Infants were able to capitalize on the added dynamic qualities of IDS to pay more attention, facilitating learning. These results illustrate how infants will readily use a wide array of information at their disposal in order to discover the structure inherent in the language in their environment.

d. Using multiple cues in speech

As researchers, we often investigate infants' abilities to perceive or use specific cues, such as phonotactics, lexical stress, sequential statistics, or pitch contour. However, infants "in the wild" are confronted with rich linguistic input that contains multiple cues to structure. Sometimes these cues are consistent with one another, which can be highly informative, while other times they may conflict, vary in their consistency, or simply provide no useful information at all. How do infants handle the complexity of natural language input, which invariably contains multiple levels of informative (and not so informative) patterns (e.g., Seidenberg, MacDonald, and Saffran, 2002)? One profitable way to study this problem is to create situations in which cues conflict. In studies of word segmentation that have used this technique, the input intentionally contains two conflicting sources of information, allowing researchers to investigate which of the two cues infants preferentially rely upon (e.g., Johnson and Jusczyk, 2001; Mattys et al., 1999; Thiessen and Saffran, 2003). These studies suggest a developmental trajectory in word segmentation: Englishlearning infants rely on different word boundary cues at different ages (whether or not this same trajectory emerges in infants learning other native languages has yet to be explored). While six-and-a-half- to seven-month-old infants appear to weight transitional probability cues over lexical stress cues (Thiessen and Saffran, 2003), eight-month-old infants weight coarticulation and stress cues over transitional probability cues (Johnson and Jusczyk, 2001), and nine-month-old infants weight stress cues over phonotactic cues (Mattys et al., 1999). This trajectory is likely a function of infants' different knowledge states at different ages. Younger infants know less than their older counterparts about how patterns like lexical stress and phonotactics work; these patterns vary from one language to the next, and require infants to know something about the structure of words in their native language. With experience, however, older infants acquire rudimentary lexical items (via sequential statistical learning, hearing words in isolation, and via segmentation of neighbors of known words such as their names). Learning new words provides infants with information about relevant patterns in their native language that may serve as cues for subsequent word segmentation (e.g., Saffran and Thiessen, 2003; Thiessen and Saffran, 2007). This emerging lexicon provides a database for the derivation of new generalizations, allowing infants to discover previously opaque regularities and patterns. In addition to presenting conflicting information, regularities in natural languages may overlap and agree much of the time.

For example, the perception of lexical stress is carried by a combination of multiple dynamic properties in the acoustic signal, including increased duration, fundamental frequency (perceived as pitch), and amplitude (loudness). While nine-monthold infants are willing to rely on any one of these properties of stress as a marker of word boundaries, older infants and adults will not (Thiessen and Saffran, 2004), suggesting that infants eventually learn how these various cues covary. The discovery of this rich correlation of acoustic information likely enhances the status of lexical stress as a word boundary cue in languages such as English. More generally, the presence and use of multiple converging cues may enhance infants' success in language learning. Computational simulations suggest that attending to multiple cues can lead to better outcomes as compared to learning via single cues (in word segmentation: Christiansen, Allen, and Seidenberg, 1998; in classification of lexical items: Reali, Christiansen, and Monaghan, 2003). Studies with infants also support the hypothesis that multiple cues can affect infants' preferences for nonsense words. In one set of studies, Saffran and Thiessen (2003) taught infants new phonotactic regularities and evaluated their ability to segment novel words that adhere to these regularities. During the first phase of the experiment, infants listened to a list of nonsense words, which all conformed to the new phonotactic regularity.

In the second phase of the experiment, infants heard a continuous stream of synthesized speech with no acoustic cues to word boundaries. Two of the words presented in the stream adhered to the phonotactic regularity, while two did not. Given this design, the words that adhered to the regularity were marked by two cues to word boundaries, transitional probabilities and the newly learned phonotactic regularities, while the boundaries of the other two words were only marked by the probability cue. In the test phase of the experiment, which used an infant-controlled preferential listening design, infants heard repetitions of all four words from the nonsense language. Infants listened significantly longer to the words that adhered to the phonotactic regularity than the words that did not. This indicates that the pattern induction phase, and possibly the presence of multiple word boundary cues, affected infants' preferences for these words. While the presence of multiple cues increased the complexity of the input, this complexity, paradoxically, likely facilitated learning. The rich structure of infant-directed speech is another example of a place where complexity may enhance learnability (e.g., Thiessen et al., 2005).

e. How does all of this happen? Learning mechanisms

Based on our brief review, which is just a sample of the burgeoning literature on infant speech and language learning (for an extensive review, see Saffran, Werker, and Werner, 2006), it is evident that discovering patterns and extracting regularities are central processes in the earliest stages of speech and language learning. As researchers, we are able to identify regularities in speech that are hypothetically useful for learning language. We then create artificial situations to allow infants to demonstrate that they can take advantage of these abilities when no other information is available. However, infants learning their native language(s) are acquiring a vastly more complex linguistic system. How do they actually go about finding the regularities that matter? These questions have led researchers to investigate the mechanisms that underlie infants' abilities to find and use regularities in linguistic input. The evidence suggests that mechanisms that support the discovery of patterns in structured input are available early in postnatal life. Kirkham, Slemmer, and Johnson (2002) demonstrated that infants as young as two months of age are sensitive to statistical patterns that occur in sequences of visual shapes, which were created to be analogous to syllable sequences. This predisposition to attend to statistical cues – a capacity that appears to be domain-general rather than specific to the domain of language (for a recent review of the issue of domain-specificity/generality in language acquisition, see Saffran and Thiessen, 2007) – may form a base from which infants can learn other cues, potentially by providing infants with a nascent corpus of words across which to discover new regularities. A recent corpus analysis by Swingley (2005) suggests that the majority of words a child would likely discover by attending to syllable co-occurrence would also reveal the predominant rhythmic pattern of the language. In this way, infants may use the statistics of syllable sequences to begin segmenting words – a cue available across languages – and then discover the language-specific and computationally less intensive strategy of using prosodic patterns to segment words.

Lending support to this hypothesis, Sahni, Saffran, and Seidenberg (2010) have shown that infants can use syllable statistics to extract an overlapping novel segmentation cue and use it to differentiate items that adhere to and violate the novel cue. They exposed ninemonth-old infants to a stream of fluent speech generated by a nonsense language containing two redundant cues to word boundaries: Words always began with /t/ (a novel cue not present in English) and contained high internal transitional probabilities (which infants of this age are able to use for word segmentation). Notably, the /t/ cue was not informative to infants prior to learning about it during the experiment. Infants next heard bisyllabic test items that either began with a /t/ or had a medial /t/ (timay vs. fotah). Infants listened significantly longer to the /t/-initial test items that adhered to the novel cue. Crucially, these items were nonwords from the nonsense language, and therefore had a transitional probability of zero. A control condition with no exposure confirmed that infants did not have a preexisting bias toward the /t/-initial items. Therefore, in order for infants to differentiate these items, they must have extracted the /t/- onset cue. Work on cue competition (e.g., Johnson and Jusczyk, 2001; Thiessen and Saffran, 2003) and cue bootstrapping (Sahni et al., 2010; Swingley, 2005; Thiessen and Saffran, 2007) implies that infants attend to and use a single reliable source of information at a time. When one cue has proven reliable, infants examine output learned via that source of information, and can then look for newly discoverable cues. However, computational models have illustrated that sequential bootstrapping of this type may be unnecessary in some cases. When simple learning mechanisms are given noisy systems containing multiple regularities at different levels of input, the system is able to capitalize on the regularities across different input units and at multiple levels. A model by Christiansen et al. (1998) shows that the presence of more regularities and patterns available in the input facilitates the model's segmentation of words from fluent speech, even if no single piece of information alone reliably points to a clear answer. Christiansen et al. (1998) trained a simple recurrent network on input from a corpus of child-directed speech.

The network was presented with words as a series of phonemes. Networks were trained on five different combinations of cues: 1) Phonological information, 2) Phonological and utterance boundary information, 3) Phonological and stress information, 4) Stress and utterance boundary information, and 5) Phonological, stress, and utterance boundary information. None of these cues independently indicated word boundaries with high reliability. The best performance was attained when the network was trained with all three cues, illustrating its ability to extract nonexplicit information from the system. It may seem obvious that the more information the network has access to, the better its performance. However, the model's performance illustrates that it is possible to extract nonexplicit information from inputs that are noisy and look confusing from the outside (like the input to child language learners). Although the model may not accurately capture how an infant receives linguistic input, it illustrates the ability to extract information using multiple probabilistic cues simultaneously. Corpus analyses support the idea that it may be beneficial for infants to use multiple segmentation cues in tandem. Curtin et al. (2005) calculated the within-word and between-word transitional probabilities in a child-directed corpus. When syllables with different stress were considered unique (i.e., an unstressed tar syllable was treated as a different syllable from a stressed tar), within-word transitional probabilities were higher than when a syllable's stress was not considered (unstressed tar and stressed tar were treated as the same syllable).

This indicates that if infants include information about stress levels in their distributional analyses, they may be more successful than when using syllable statistics alone. When focusing on learning mechanisms, it is important to be clear that invoking learning does not imply the presence of a blank slate: all environment, with no internal structure. Indeed, the evidence strongly suggests that constraints on learning play a central role in language acquisition; all patterns are not equally learnable (e.g., Saffran, 2003). For example, Newport and Aslin (2004) demonstrated that when adult learners are exposed to sound sequences containing nonadjacent regularities (e.g., AXB, where X varies such that the relevant dependency is between A and B), some such sequences are more learnable than others. When A, X, and B are individual segments, and A and B are similar in kind (A and B are consonants and X is a vowel, or vice versa), adults succeed at discovering the relevant dependencies. However, when A, X, and B are all syllables, adults fail to find the relevant dependency. These results extend to domains beyond language (Creel, Newport, and Aslin, 2004), and suggest that perceptual constraints on grouping, as suggested long ago by Gestalt psychologists, affect sequence learning. Notably, the same constraints on learning were not observed in nonhuman primates learning the same sequences (Newport et al., 2004). The kinds of patterns most readily learned by humans are those that also are most likely to occur in human languages: for example, phonological systems that include dependencies between consonants that span vowels or vice versa (as in languages that make use of consonant or vowel harmony), but not dependencies between syllables that span syllables. This relationship between learnability and presence in human languages may be nonaccidental; the ease with which humans can learn a particular structure may influence the likelihood that that structure occurs cross-linguistically.

f. Extracting regularities: The gateway to language

The ability to extract regularities over multiple units of sound is extremely helpful when finding and recognizing words. However, there is much more to knowing language than just being able to recognize words. Abilities to recognize and use patterns, in the form of learning mechanisms that extract structure via regularity detection, may be the gateway to language. For example, infants can track patterns of words to find grammar-like sequences, and recognize those same sequences when they are exemplified with novel words (e.g., Gómez and Gerken, 1999; Marcus et al., 1999). Under some circumstances, infants can discover nonadjacent dependencies between elements in sequence, suggesting the availability of learning mechanisms not tightly tied to sequential order (e.g., Gómez, 2002). Infants can even perform word-sequence learning tasks given unsegmented fluent speech: When twelve-month-old infants are exposed to a sequence of syllables organized into words, and words organized into sentences, they can discover and learn both of these layers of patterns superimposed in the same set of stimuli, as in natural languages (Saffran and Wilson, 2003). It is also clear that early learning about the sound patterns of language influences later language learning. Infants' skill at word segmentation is a predictor of later success at word learning in the native language (Newman et al., 2006). When engaged in word segmentation, even in a novel nonsense language (with English phonology), infants appear to treat the output of the segmentation process as candidate novel words in English (Curtin et al., 2005; Saffran, 2001). Many other features of the sounds of novel words influence the ease with which infants can map them to meanings (for a recent review, see Saffran and Graf Estes, 2006). Learning structures at one level of language – even the lowest levels of sound – has implications for many other language learning problems, including grammatical categories and aspects of syntax (e.g., Kelly and Martin, 1994). Moreover, to the extent that learning is linked to language processing (Seidenberg and MacDonald, 1999), we would expect to see that the factors influencing learning also influence later language processing.

g. What does the future hold?

Modern developmental scientists are focused on investigating questions of how and why, rather than what or when. In order to pursue these deeper questions of mechanism, researchers have begun to harness the power of applying multiple methodologies to individual problems in the development of speech and early language (e.g., Hollich, 2006; Kuhl, 2004). Along with more sophisticated scientific questions, the field requires more sophisticated testing methods. One technological advance has been the use of eyetracking methods in infant speech and language studies (e.g., Aslin and McMurray, 2004; Fernald, Swingley, and Pinto, 2001). Traditionally, researchers have used infants' head movements to assess whether infants are attending to one stimulus over another. However, by using eyetracking methods, we can more accurately assess which stimulus the infant is attending to, as well as the details of infants' phonological and lexical representations. McMurray and Aslin (2004) developed a forced choice paradigm that capitalizes on this technology. In this task, infants concurrently saw a looming circle on a video monitor and heard one of two words, for example, lamb or teak. Next, the circle moved behind a T-shaped occluder and emerged either on the top right or top left of the T. If the circle was paired with lamb it emerged on the top left. If it was paired with teak it emerged on the top right. After training, infants were tested to see if they could generalize to other exemplars of lamb and teak in which duration and pitch were varied. While infants could generalize over different pitch ranges, they had difficulty generalizing over longer durations. By using this forced choice eyetracking method, McMurray and Aslin were able to tap a response that is behaviorally less taxing and capitalizes on infants' tendency to track moving objects. The reduction of task demands is well-known to radically alter the kinds of knowledge that infants can demonstrate in an experiment (for demonstrations elsewhere in infant cognition, see Keen, 2003; Munakata, et al., 1997). It is likely that continued advances in behavioral techniques will permit ever more detailed understanding of infants' internal representations and learning processes. Noninvasive neuroimaging methods, including fMRI, MEG, and optical imaging, along with the more traditionally used EEG and ERP, are also likely to help us to uncover the underpinnings of language acquisition (e.g., Dehaene-Lambertz et al., 2002; Peña et al., 2003). While these methods, to date, have largely been used to ask "where" and "what" questions, they may be profitably applied to the study of "how" by investigating the circuitry and domain-specificity/generality of the mechanisms that subserve language learning. These methods have been effectively employed in studies of adults that have moved the field beyond old-fashioned general statements concerning hemispheric asymmetries (e.g., "language is localized in the left hemisphere, with music on the right") to more nuanced theoretical constructs (e.g., Scott, this volume; Zatorre and Belin, 2001). Another innovation that allows researchers to ask deeper questions is the use of crossspecies comparisons (Hauser, Chomsky, and Fitch, 2002). Humans are the only species that possesses a full linguistic system. What mechanisms or abilities do humans possess that allow them to have this unique skill? By performing cross-species comparisons we may move closer to answering these questions. For example, the ability of infants to use probability and frequency information to segment words from fluent speech seems to be a powerful mechanism (e.g., Saffran et al., 1996). This may be a crucial early step in the acquisition of language. However, Hauser, Newport, and Aslin (2001) found that adult cotton top tamarins are also able to capitalize on the same frequency and probability information. The fact that both tamarins and human infants are able to use this type of regularity suggests that there must be a more sophisticated skill that is a point of divergence between the two species. By comparing what humans and nonhuman primates can do, we can understand more about the evolution of language. Further work on more sophisticated skills, such as learning nonadjacent dependencies between speech sounds (e.g., Newport et al., 2004) and learning complex grammars (e.g., Saffran et al., 2008) have illuminated points of divergence between the species. By examining the difference in skills between nonhuman primates and humans we can better understand the “how” and “why” questions of language acquisition. In conclusion, we have seen a great upsurge in research focused on the ontogenesis of speech and language in infancy. It is now evident that however much the infant brings to this task in terms of factory-installed predispositions and perceptual systems, it is our learning mechanisms, operating under the constraints placed by our perceptual and cognitive systems, which become critically important early in postnatal life. There is still much about this process that we do not understand; most notably, the stimuli used in these experiments vastly underestimate the complexity of the problems facing the language learning child. However, increasingly sophisticated methodology in tandem with increasingly rich theorizing are continuously moving researchers closer to understanding the unfolding of infant speech perception and its relationship to the beginnings of language learning.

2. Gestural phonology

The tenets of gestural phonology are grounded in the spatiotemporal organization of articulatory gestures in speech, which are themselves grounded in the biomechanical organization ofthe human vocal tract. Rather than assuming abstract and timeless phonetic features as the atoms or primitives from which phonological representations are built, the gestural model assumes that the phonological primitives are articulatory gestures, the coordinated actions of vocal tract articulators. The model organizes these gestural features within the framework of a hierarchical articulatory geometry based on the anatomical relations among the articulators involved in speech. The vocal tract is comprised of three relatively independent articulatory systems that are represented as separate nodes within the articulatory geometry: the glottal system (vocal cords), the nasal system (the velum, the valve that permits or prohibits air flow through the nasal cavity), and the oral system, which includes the lips and the tongue as separate subsystems. There is an additional subordinate level in the tongue subsystem: tongue tip Versus tongue body, whose actions are differentiated by different intrinsic and extrinsic muscles of the tongue. This hierarchically organized set of articulators functions within the confines ofthe walls ofthe vocal tract, which is structured basically as a bent tube of varying diameter, optionally connected to a second side tube (nasal cavity) via the open velum. The coordinated actions of the articulators can cause constrictions at various locations (place of articulation) along the vocal tract (e.g., dental, alveolar, velar, etc.) (see Figure 1 for additional places of articulation). Each place can display several variations in degree of constriction, which determines the manner of the sound produced (complete closure for stop consonants, critical constriction for causing turbulent airflow in fricatives, narrow constriction for some vowels and for approximant consonants such as Iwl and Ir/, wide opening for the velum in nasals and the glottis in voiceless sounds).

Articulatory geometry is compatible, in many respects, the with nonlinear or autosegmental approaches that have supplanted SPE phonology. Some important distinctions must be noted, however, between the two approaches. Specifically, gestural phonology posits phonological elements to be gestures defined by a set of dynamic equations describing the movement of articulators over space and time, rather than a specification of abstract, timeless phonetic features. To illustrate, the equation set for the syllable ma describes a velum opening gesture and lip closing gesture which begin simultaneously and reach their peaks synchronously to produce the Im/, and a slower, less extreme tongue body gesture to narrow the pharynx (upper throat) for the "ah" vowel, which begins synchronous with the other two gestures but peaks later and lasts longer. Thus, articulatory geometry is closely related to the anatomical structures and movement patterns of the vocal tract. This way, in the gestural model the phonological primitives and their physical instantiations derive from a single domain grounded in the spatiotemporal properties of real articulatory events. Because of this, phonological representations can specifY the relative timing, or phasing, of one articulatory gesture relative to another. For example, the Canadian French versus continental French difference in vowel nasalization that was mentioned earlier (van Reenin, 1982) can be specified dynamically as a difference in the relative timing, or phasing, between the onset of velum lowering for nazalization and the peak of tongue movement for the vowel. This characterization departs critically from the phonetics-phonology relationship held by classic SPE phonology and by nonlinear phonologies, neither of which can phonologically represent the dialectal difference phonologically, even though the nasalization difference appears to be part of the language-specific grammar in the two dialects. This representational inability occurs for the latter two views because they posit that phonetic and phonological information exists in two divergent, informationally incompatible domains, one physical (actual articulations) and the other only mental (underlying phonological representations).

3. Language-specific phonetic-gestural properties and perceptual learning

Properties and perceptual learning Recall the basic tenets of perceptual learning according to the ecological perspective~that perceptual systems become attuned by experience to particular types ofinformation; that this involves optimization in the pickup of relevant information; that it entails the discovery of critically distinguishing properties of distal structures and events; and that this is accomplished via perceivers' active search for invariants in the flow of stimulation that most economically specify those crucial properties. Educated attention minimizes uncertainty about objects and events in the world, by selecting or extracting reduced information specifically for its ability to critically differentiate things of interest or usefulness to the perceiver. Earlier it was argued that the identity of objects and events is specified by structural and transformational invariants available in the flow of stimulation over time and space. Moreover, recognition ofsimilarities and differences among things often depends on abstraction of higher-order invariants which depend on prior detection of other, lower-order invariants. As Eleanor Gibson remarked, the critical invariants are generally relational in nature, rather than isolated, independent attributes. To consider how higher-order relational invariants might be discovered in speech through perceptual learning, I will turn briefly to some central concepts developed in work on an ecological approach to the formation of complex coordinated skills and behaviors (e.g., Kugler, Kelso, & Turvey, 1982; Saltzman & Kelso, 1987; Turvey, 1980; 1990) including speech (Saltzman & Munhall, 1989). The goal of coordination is to maximize the adaptability and flexibility of achieving some goal of action by minimizing the number of separate dimensions that must be directly controlled. As Turvey (e.g., 1980, 1990) and others have argued, this is accomplished by forming task-specific synergies among muscle groups, or cQordinative structures. To understand this concept, consider an example commonly cited by ecological researchers-the task of a puppeteer and the way that the construction of her marionette simplifies the control of its movements. By linking the puppet's limbs with strings to a controller bar, the puppeteer obviates the need to move each joint of each limb separately, instead producing coordinated movements among multiple limbs by single movement of the controller. By this means, the many degrees of freedom controlling the joints of the separate limbs have become joined together into a coordinative structure with fewer degrees that must be directly controlled. Research on locomotion indicates that coordinative structures account for the coordination offlex- ion and extension of each leg joint in proper sequence during the swing of each leg, the alternation between the legs, and the postural adjustments required throughout for maintenance of balance. Coordinative structures show task-specific flexibility in that temporary perturbations result in automatic, immediate compensatory adjustments among the coordinated elements so that the general goal is preserved without requiring numerous command decisions about specific elements.

Saltzman and Munhall (1989) provide logical and empirical evidence that in speech coordinative structures accomplish the gestural goal offorming a constriction of a particular degree at a particular vocal tract location, by harnessing together the specific articulators in ways that automatically compensate for perturbations and contextual variations. The language-specific gestural phasing patterns of Browman and Goldstein's gestural constellations are examples of higher-order coordinative structures in speech. Coordinative structures in motor control can form and re-form, and operate as emergent properties of selforganizing systems (see Madore & Freeman, 1987; Prigogine, 1980; Prigogine & Stengers, 1984; SchOner & Kelso, 1988; Turvey, 1980, 1990). Emergent properties of self-organizing systems, including their sensitivity to initial conditions, have been proposed as the basis for the evolution of maximal dispersion among the elements of language-specific phonological inventories (Lindblom, 1992; Lindblom, Krull, & Stark, 1993; Lindblom, MacNeilage, & Studdert-Kennedy, 1983), as well as for the ontogeny of phonological organization in the child (Mohanan, 1992; Studdert-Kennedy, 1989). The latter proposals point to the importance of viewing the native phonology as an organized system when considering how language-specific experience may affect perception of phonetic patterns that fall outside the native phonological system.

CONCLUSION

What is innate about the development of the phonological component of a language's grammar? That is, what is it that provides the constraints on acquisition of possible phonological systems? By the ecological reasoning presented in this chapter, the answer is that what is innate-what provides the constraints on phonologies and their development-is the structure and dynamic possibilities of the human vocal tract. To a first approximation, this claim is in line with the underlying assumptions of Chomsky and Halle themselves, whose universal phonetic features were initially based on articulatory concepts. The point on which I disagree with them is their assumption that the constraints are specified innately in the mind. By the ecological view proposed here, the constraints are, instead, literally in the physical head, in the vocal tract itself and in the lawful physical effects that its configuration and movements have on the temporally-varying shape ofits acoustic product. Chomsky and Halle (1968) were correct in suggesting that the listener who knows a language hears the phonetic shapes made familiar by experience with that language. This claim, I have argued, can be extended even to predict that the listener hears echoes of those familiar, native phonetic shapes in the non-native sounds and contrasts of unfamiliar languages. But I part ways with their reasoning about the causal mechanisms, and about the source of listeners' knowledge. Instead, I claim that listeners hear the phonological structure of their native language in non-native speech because they have learned to detect the gestural invariants that are directly available in the information flow from the language environment. Listeners become attuned to these gestural patterns and pick up the invariants specifying those familiar patterns wherever the stimulation provides criterial evidence for them, even in non-native sounds. This attunement to native gestural invariants begins in infancy but extends over development and into adulthood, where it should even help to account for perceptual changes during the learning of additional languages.

We have seen that the mere phonetic framework of speech does not constitute the inner fact of language and that the single sound of articulated speech is not, as such, a linguistic element at all. For all that, speech is so inevitably bound up with sounds and their articulation that we can hardly avoid giving the subject of phonetics some general consideration. Experience has shown that neither the purely formal aspects of a language nor the course of its history can be fully understood without reference to the sounds in which this form and this history are embodied. A detailed survey of phonetics would be both too technical for the general reader and too loosely related to our main theme to warrant the needed space, but we can well afford to consider a few outstanding facts and ideas connected with the sounds of language. The feeling that the average speaker has of his language is that it is built up, acoustically speaking, of a comparatively small number of distinct sounds, each of which is rather accurately provided for in the current alphabet by one letter or, in a few cases, by two or more alternative letters. As for the languages of foreigners, he generally feels that, aside from a few striking differences that cannot escape even the uncritical ear, the sounds they use are the same as those he is familiar with but that there is a mysterious "accent" to these foreign languages, a certain unanalyzed phonetic character, apart from the sounds as such, that gives them their air of strangeness. This naïve feeling is largely illusory on both scores. Phonetic analysis convinces one that the number of clearly distinguishable sounds and nuances of sounds that are habitually employed by the speakers of a language is far greater than they themselves recognize. Probably not one English speaker out of a hundred has the remotest idea that the *t* of a word like *sting* is not at all the same sound as the *t* of *teem,* the latter *t* having a fullness of "breath release" that is inhibited in the former case by the preceding *s;* that the *ea* of *meat* is of perceptibly shorter duration than the *ea* of *mead;* or that the final *s* of a word like *heads* is not the full, buzzing *z* sound of the *s* in such a word as *please.*

It is the frequent failure of foreigners, who have acquired a practical mastery of English and who have eliminated all the cruder phonetic shortcomings of their less careful brethren, to observe such minor distinctions that helps to give their English pronunciation the curiously elusive "accent" that we all vaguely feel. We do not diagnose the "accent" as the total acoustic effect produced by a series of slight but specific phonetic errors for the very good reason that we have never made clear to ourselves our own phonetic stock in trade. If two languages taken at random, say English and Russian, are compared as to their phonetic systems, we are more apt than not to find that very few of the phonetic elements of the one find an exact analogue in the other. Thus, the *t* of a Russian word like *tam* "there" is neither the

English *t* of *sting* nor the English *t* of *teem.* It differs from both in its "dental" articulation, in other words, in being produced by contact of the tip of the tongue with the upper teeth, not, as in English, by contact of the tongue back of the tip with the gum ridge above the teeth; moreover, it differs from the *t* of *teem* also in the absence of a marked "breath release" before the following vowel is attached, so that its acoustic effect is of a more precise, "metallic" nature than in English. Again, the English *l* is unknown in Russian, which possesses, on the other hand, two distinct *l*-sounds that the normal English speaker would find it difficult exactly to reproduce—a "hollow," guttural-like *l* and a "soft," palatalized *l*-sound that is only very approximately rendered, in English terms, as *ly*. Even so simple and, one would imagine, so invariable a sound as *m* differs in the two languages. In a Russian word like *most* "bridge" the *m* is not the same as the *m* of the English word *most;* the lips are more fully rounded during its articulation, so that it makes a heavier, more resonant impression on the ear. The vowels, needless to say, differ completely in English and Russian, hardly any two of them being quite the same.

[This Page Is Intentionally Left Blank]

CHAPTER 5
THE PERCEPTION OF SPEECH

INTRODUCTION

During the second half of the 20th century, research concerning speech perception stood relatively distinct from the study of audition and other modalities of high-level perception such as vision. Contemporary research, however, is beginning to bridge this traditional divide. Fundamental principles that govern all perception, some known for more than a century, are shaping our understanding of perception of speech as well as other familiar sounds.

Investigators of speech perception traditionally attempted to explain how listeners perceive the spoken acoustic signal as a sequence of consonants and vowels, collectively referred to as phonetic segments or units. When one describes speech sounds in this way, brackets are used to surround phonetic symbols such as [j] (the ‘y’ sound in ‘yes’) and [o] (as in ‘oh’). By contrast, phonemes are more abstract linguistic units that roughly correspond to letters in written language, and are transcribed surrounded by slashes (/j/ and /o/.) Morphemes are the smallest meaningful units of language, roughly corresponding to words (e.g., ‘dog’, ‘taste’, as well as ‘dis’- and -‘ful’) with phonemes being the smallest units that can change the meaning of a morpheme (e.g., ‘yo’ versus ‘go’) (Trubetskoy, 1969). Within this scheme, the experimental study of speech perception classically has corresponded more or less to the lowest division of labor generally agreed upon by linguists and psycholinguists.

To the extent that speech perception researchers’ task is to deliver minimal units to those who study language, an important caveat must be applied to this inherited division of labor. There is no clear experimental evidence demonstrating that either phonetic segments or phonemes are real outside of linguistic theory (e.g., Lotto, 2000), and the appeal of phonetic segments and phonemes may arise principally from experience with alphabetic writing systems (e.g., Morais, Bertelson, Cary, & Alegria, 1986; Morais, Cary, Alegria, & Bertelson, 1979; Port, in press). One ought not be sanguine about whether speech perception really is about recognizing consonants and vowels per se. Listeners probably do not extract phonemes preliminary to recognizing words. There may be nowhere in the brain where phonemes reside independent of words that they comprise.

Nevertheless, conceptualizing speech perception as a process by which phonemes are retrieved from acoustic signals is tradition. Within this tradition, research in speech perception often has been focused on problems concerning segmentation and lack of invariance. The problem of segmentation refers to the fact that, if phonetic units exist, they are not like typed letters on a page. Instead, they overlap extensively in time, much like cursive handwriting. The problem of lack of invariance (or, problem of variability) is related to the segmentation problem. Because speech sounds are produced such that articulations for one consonant or vowel overlaps with production of preceding ones, and vice versa, every consonant and vowel produced in fluent connected speech is dramatically colored by its neighbors. Some of the most recalcitrant problems in the study of speech perception are the consequence of adopting discrete phonetic units as a level of analysis, a level that is not discrete and may not be real. In connected speech, acoustic realization of the beginning and end of one word also overlaps with sounds of preceding and following words, so the problems of invariance and segmentation are not restricted to phonetic units.

This being said, either morphemes or words are the first units of language that stand more or less on their own accord.1 It is possible, even likely, that speech perception is a series of non-discrete processes along the way from waveforms to words. In this chapter, speech perception will be described as a continuum of processes operating on the acoustic signal with varying levels of sophistication. The consistent theme will be common principles that define how these processes work.

Following some preliminaries concerning broad principles that govern perception, a framework for conceptualizing perception of speech will be presented. In part, this approach is modest because many of the central premises are derivative of what is known about domain-general processes of perception and learning. In addition, this approach is conservative by virtue of avoiding ad hoc claims concerning processing of speech in any unique way, while also avoiding reliance upon higher-level cognitive processes. The central claim is that perception of speech works the same way perception works for other modalities and for other environmental sources. Speech perception follows a handful of general principles that are implemented in both sophisticated and not-so-sophisticated ways through the chain of processing from periphery through central nervous system.

A. Speech Perception

Before the advent of modern signal processing technology, linguists and psychologists believed that speech perception was a fairly uncomplicated, straightforward process. Theoretical linguistics' description of spoken language relied on the use of sequential strings of abstract, context-invariant segments, or phonemes, which provided the mechanism of contrast between lexical items (e.g., distinguishing pat from bat).1,2 The immense analytic success and relative ease of approaches using such symbolic structures led language researchers to believe that the physical implementation of speech would adhere to the segmental 'linearity condition,' so that the acoustics corresponding to consecutive phonemes would concatenate like an acoustic alphabet or a string of beads stretched out in time. If that were the case, perception of the linguistic message in spoken utterances would be a trivial matching process of acoustics to contrastive phonemes.3 Understanding the true nature of the physical speech signal, however, has turned out to be far from easy. Early signal processing technologies, prior to the 1940s, could detect and display time-varying acoustic amplitudes in speech, resulting in the familiar waveform seen in Figure 1. Phoneticians have long known that it is the component frequencies encoded within speech acoustics, and how they vary over time, that serve to distinguish one speech percept from another, but waveforms do not readily provide access to this key information. A major breakthrough came in 1946, when Ralph Potter and his colleagues at Bell Laboratories developed the speech spectrogram, a representation which uses the mathematical Fourier transform to uncover the strength of the speech signal hidden in the waveform amplitudes (as shown in Figure 1) at a wide range of possible component frequencies.4 Each calculation finds the signal strength through the frequency spectrum of a small time window of the speech waveform; stringing the results of these time-window analyses together yields a speech spectrogram or voiceprint, representing the dynamic frequency characteristics of the spoken signal as it changes over time (Figure 2).

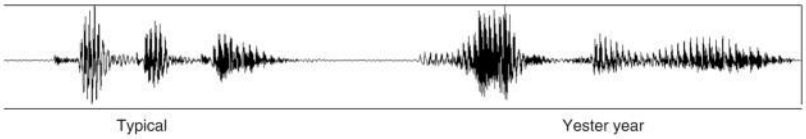


Figure 1

Speech waveform of the words typical and yesteryear as produced by an adult male speaker, representing variations in amplitude over time. Vowels are generally the most resonant speech component, corresponding to the most extreme amplitude levels seen here. The identifying formant frequency information in the acoustics is not readily accessible from visual inspection of waveforms such as these.

Figure 2

A wide-band speech spectrogram of the same utterance as in Figure 1, showing the change in component frequencies over time. Frequency is represented along the y-axis and time on the x-axis. Darkness corresponds to greater signal strength at the corresponding frequency and time. A talker can expect a listener to grasp the rough dimension of any sincere and appropriate message, though only by saying it. For talker and listener, speech is a medium, a link in a commonplace causal chain by which pleasantries or philosophies are exchanged, cooperation is negotiated and compliance is compelled. But, does an essay about speech belong in a book about language? To a newcomer, it is self-evident that conversational partners know what each other says simply by hearing the sounds of spoken words. From this perspective, the fundamentals of speech perception surely lie in psychoacoustics, an essential reduction of speech perception to sensory resolution and auditory categorization. Even so, the newcomer might already notice the difference in auditory quality in the speech of children and adults, or in face to face and in telephone speech, and suspect that the perception of spoken messages entails more than acute hearing. To the old hand familiar with cognitive psychology and the historic place of speech within it, the motivation to study speech perception might seem well and truly relieved now that affordable devices transcribe words from sound. On the contrary, this essay like its companions in this volume was produced by a typing hand and not by a dictating voice, despite the mathematical ingenuity of the engineers – far exceeding that of cognitive psychologists – who create speech-to-text devices. For the reader of any degree of experience, this part of the Handbook explains why the descriptive and theoretical puzzles provoked by speech perception have proven to be so enduring, psychologically and linguistically, and in doing so claims a role for speech in language.

Our characterization of the perception of speech ranges across three of its facets. First, we discuss the historic aim of research on speech, which has been to understand how acoustic properties evoke an impression of linguistic form. This line of research is mature, and a sizeable literature beginning with classical sources presents a consistent expression of competing views and evidence. Ideological commitments aside, it is a singular merit of this research tradition that it introduced a generous assortment of theoretical conceptualizations to perceptual psychology. Even when innovation happened to spring from other sources, the well established techniques and research paradigms within the study of speech perception provided a ready means to calibrate the explanatory adequacy of a principle. This portion of the essay exposes contemporary viewpoints about perceptual organization and analysis of speech and notes the questions that lead the research forward.

Second, the ordinary perceptual resolution of the linguistic properties of speech is accompanied by an irreducible impression of the talker as well as the message. Research about the recognition of individuals from their speech takes its origin in forensic projects – studies to determine whether a known talker and an unidentified talker are the same – and in artifactual methods to create a vocal identification technology. In contrast to these humble roots, more recent cognitive studies emphasize the perceptual effects of variation in phonetic form across individuals and instances. The evident perceptual interchange of linguistic, individual (or, indexical) and situated properties promised to overturn the classic conceptualization of the acoustic-to-phonetic projection, and this portion of our essay describes the partial success of this project and the questions that remain for a complete causal account.

The third section of our essay characterizes self-regulatory speech perception in which an individual talker's self-perception modulates the production of speech. This theme is contrary to Lashley's founding arguments in psycholinguistics. He held that the rate of production of vocal actions was too rapid to permit monitoring by proprioception, and many studies since have recounted adequate unmonitored articulation, for instance, concurrent to mandibular somatic sensory blockade. This literature about the control of coordination in vocal movement is supplemented and elaborated by more recent studies that identified effects of self-monitoring in other sensory modalities. These findings show that talkers adjust subtle – and, less subtle – properties of articulatory expression as a consequence of phonetic perception, albeit at a slower pace than Lashley stipulated, and in varied social conditions.

Throughout, our essay is organized by psycholinguistic questions, rather than by concerns with specific research methods. Although the investigations that we describe are largely the yield of functional studies of normal adults, we have referred to research about special populations or using special methods when we aimed to secure premises in our argument. We also direct the reader to other discussions when technical matters or special perceivers hold intrinsic interest or importance.

1. Perceptual Organization and Analysis of Speech

A listener intent on grasping a talker's message must sample physical effects of speech that vary regularly if unpredictably, a consequence of a talker's vocal acts. The regularity as well as the unpredictability derive from a common cause; the linguistic governance of speech deploys formal attributes designated in the talker's language, and these drive the regularities. At the same time, no expression is an exact repetition of a prior one, and whether the departures from stereotypy are attributed to chance or to a specific cause – to a talker's enthusiasm, or haste, or influenza – exact patterns never recur. The central problem in research on speech has been to understand how perception of regular linguistic attributes is evoked by such unpredictably varying acoustic causes.

None of the acoustic constituents of speech is unique to speech, although some features of speech are characteristic: a cyclical rise and fall of energy associated with a train of syllables, amplitude peaks and valleys in the short-term spectrum, and variation over time in the frequency at which the peaks and valleys occur (Stevens & Blumstein, 1981). In addition to noting these attributes, it is fair to say that natural speech is an acoustic composite of whistles, clicks, hisses, buzzes and hums, a discontinuous and often aperiodic result of the continuous movement of articulators. In following a speech signal, a listener tracks an intermittent pattern of heterogeneous acoustic constituents; there is no single element nor set of them that defines speech, therefore, no simple way for a perceiver to distinguish speech piecemeal from the acoustic effects of other sources of sound. Despite all, a perceiver often tracks the speech of a specific talker sampling by ear and eye, two kinds of perceptual organization that also combine multi-modally, and resolves the linguistic properties in the sensory effects – that is to say, perceptual analysis of the symbolic properties of speech succeeds. We discuss these in turn.

a. Perceptual Organization

The ability to track an individual's speech amid other sounds retains the characterization applied long ago by Cherry (1953), the cocktail party problem. Such get-togethers can pose many challenges for participants; this specific cocktail party problem is solved by perceivers who understand spoken messages despite the concurrent intrusions of acoustic elements very much like those composing the target speech stream. The sources of unrelated sounds surely include the clinking of glasses and popping of corks, although other extraneous acoustic moments are similar to an attended speech stream because they come from the speech of other talkers. Indoors, the direct sound mixes with late arriving reflections from the ceiling, floor and walls of the attended speech signal itself.

To gauge the means of resolving the sound produced by a single individual, the contrast between visual and auditory attention is instructive. In attending to a visible object or event, a perceiver typically turns to face it, bringing the light reflected by the object of interest to the fovea of the retina. In this retinal region, receptors are densest and pattern acuity is best, for which reasons visual attention will often coincide with a foveated object. A listener's attention to the audible world achieves spatial and spectral focus psychologically, without the selective benefit of a heading at which auditory pattern acuity peaks. In addition, the visible world contains opaque, translucent and transparent objects; the audible world is largely transparent. A listener cannot presume that a sound arriving from a certain direction stems from the visible object at the same heading, for sounds produced by other sources at the same direction are likely to propagate around intermediate objects to impinge on the listener. Despite all, perception often reciprocates the patterned variation of a speech stream with its discontinuities, dissimilarities among components and similarities between its components and those of unattended utterances and other events. This perceptual function is fast, unlearned, keyed to complex patterns of sensory variation, tolerant of anomalous sensory quality, nonsymbolic and dependent on attention whether elicited or exerted (Remez, Rubin, Berns, Pardo, & Lang, 1994). The evidence to characterize the function and the limits of its effectiveness stems from several lines of research.

1) Fast

Whether speech occurs in the clear or in noise, it is quickly resolved perceptually if it is resolved at all. Classic studies of the persistence of the auditory trace of speech indicate such fast resolution, for they show that discrimination based on an auditory form of speech becomes poor very rapidly. Before the sensory trace fades, the auditory effects of speech are resolved into a coherent perceptual stream. The estimates of the rate of decay vary, though we can be certain that little of the raw auditory impression of speech is available after 100 ms (Elliott, 1962); and, none after 400 ms (Howell & Darwin, 1977; Pisoni & Tash, 1974). For the perceiver, the perishable auditory form creates an urgent limit on integration of the diverse constituents of speech; auditory properties available to perception are simply lost if integration is delayed. For a theorist, the evident long-term adaptive flexibility exhibited in natural perception cannot be attributed to unelaborated representations of the auditory features of speech without denying this basic psychoacoustic limit (see Grossberg, 2003). In contrast to the natural perceiver, urgency does not constrain artefactual recognizers. The schemes that they employ inherently surpass the physiological characteristics of an auditory system. They can sample and hold acoustic representations of speech analogous to the initial auditory sensory forms; indeed, they can hold them as long as electricity powers the memory (Klatt, 1989). Such superhuman systems have had wide theoretical influence despite indifference to the critical first step of urgent perceptual organization (Picheny, 2003).

2) Unlearned

Evidence that perceptual organization of speech is unlearned derives from studies of 14 week old infants, who integrated acoustic elements of speech composed through synthesis to be both spectrally and spatially disparate (Eimas & Miller, 1992; cf. Hollich, Newman, & Jusczyk, 2005). Listeners at this young age are hardly aware of linguistic properties in the speech they apprehend, and the perceptual coherence of the diverse constituents can be attributed to precocious sensitivity to vocalization independent of phonetic impressions, and well in advance of linguistic sensitivity. If experience plays a bootstrapping role in perceptual organization during the first three months of life, this is unlikely to entail arduous tutelage, nor sleep learning via exposure to adults whispering in the nursery.

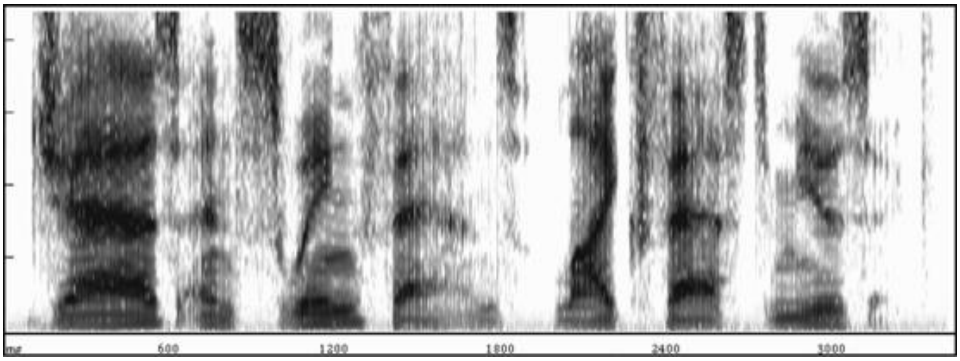
3) Keyed to complex patterns of sensory variation.

The amplitude peaks and valleys in the spectrum of speech are natural resonances of the column of air enclosed within the anatomy of the upper airway. These resonances, or formants, are set ringing by the regular pulsing of the larynx, which produces harmonic excitation; or, by the production and release of air pressure behind an approximation or occlusion, as in the case of stop consonants; or, by sustained turbulence, as in the case of frication and aspiration. Acoustic changes in the spectrum are nonuniform across the formants. Specifically, the independent control of the articulators that produces formant frequency variation causes uncorrelated differences across the formants in the extent, rise and fall and temporal relation of frequency and amplitude change. Equal change in the first, second, third, nasal and fricative formants is uncharacteristic of vocal sound production, and aggregation of the sensory correlates of speech in perceptual organization occurs without evident reliance on similarity of change across the resonances. In some acoustic transforms of speech spectra, the frequency variation of the resonances is obscured without loss of perceptual coherence. In one version aiming to model the diminished frequency resolution imposed by an electrocochlear prosthesis (Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995) the coarse shape of the short-term spectrum envelope was represented in the power of 3 or 4 noise bands, each spanning a large portion of the frequencies of speech. Over time, asynchronous amplitude variation across the noise bands creates a derivative of speech without harmonic excitation and broadband formants, yet the frequency contours of individual resonances are absent. The effectiveness of such variants of speech spectra exposes the basis for the perceptual organization of speech, which lies in detecting a sensory pattern that coincides with phonologically governed articulation. Although these remain to be characterized formally, to a first approximation it is clear that the patterns of sensory variation are complex.

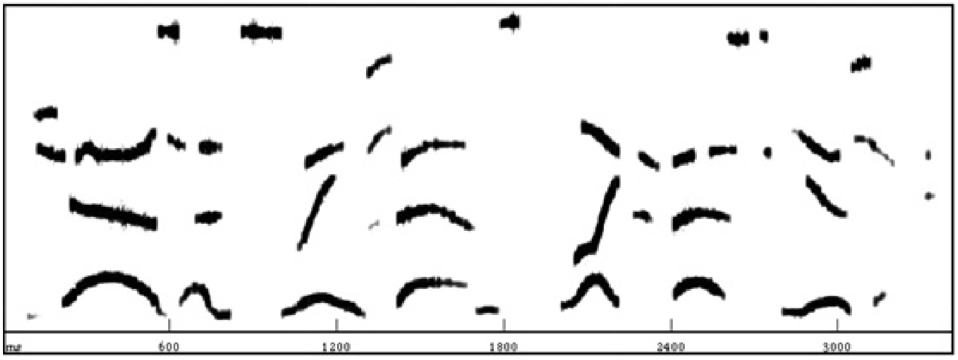
4) Tolerance of anomalous auditory quality

Perceptual organization of speech is tolerant of anomalous auditory quality, as research with sinewave replicas of utterances has shown (Remez, Rubin, Pisoni, & Carrell, 1981). In this synthetic acoustic signal, the natural products of vocalization are eliminated by imposing the pattern of a speech spectrum on elements that are not vocal in origin. Precisely, three or four pure tones are set to vary in frequency and amplitude in the pattern of the estimated formant peaks of a speech sample. A fourth tone is intermittently used to replicate fricative formants, brief bursts or nasal murmurs. The integration of the tones to compose an intelligible utterance occurs despite the persistence of the weird quality of a sinewave voice, evidence that neither natural acoustic correlates of speech nor auditory impressions of a legitimate voice are required for perceptual organization to occur. Studies with chimerical signals provide independent corroboration that the perceptual organization of speech is indifferent to the specific acoustic constituents of a signal and to the nonvocal auditory quality that can result (Smith, Delgutte, & Oxenham, 2002). To create an acoustic chimera, a coarse grain representation of the spectrum envelope of speech is excited with an arbitrarily chosen source. The result is a composite exhibiting the influence of each aspect, the spoken utterance and the arbitrary source. Like tone analogs of speech, a chimera is intelligible linguistically, evidence that its constituents are grouped to compose a signal fit to analyze as speech. Phenomenally, it retains the quality of the excitation, whether noisy, or harmonic, or, indeed, multiple, as in the instance shown in Figure 1, for which the excitation was provided from an acoustic sample of a musical ensemble. In each of these critical cases, the perceptual coherence survived the inventory of arbitrary and nonvocal short-term properties by tracking the time-varying properties, which derived from speech even when the acoustic elements did not. This series of findings eliminates as implausible any characterization of perception warranting meticulous attention to elemental speech sounds or their correlated qualitative effects. Instead, they make evident the causal properties of a spectrotemporal pattern superordinate to the momentary constituents.

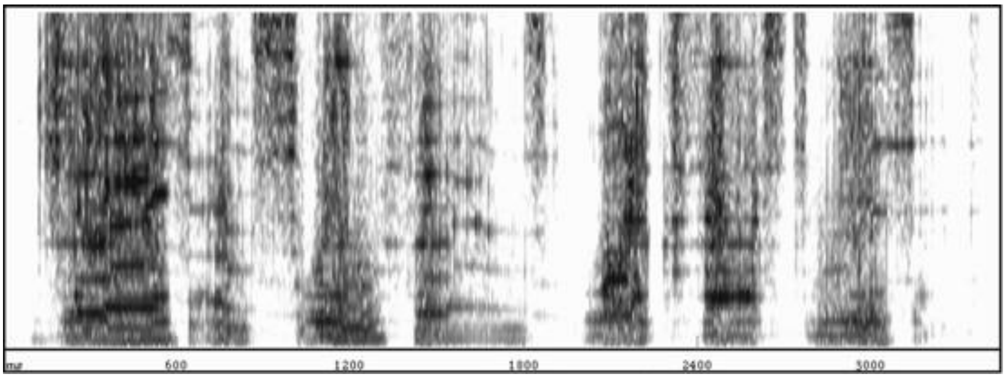
A. Natural



B. Sinewave replica



C. Chimera of speech and music



Inventory of acoustic constituents of three kinds of intelligible sentence: (A) natural speech; (B) synthetic replica composed of time-varying sinusoids; and (C) acoustic chimera made by exciting the changing spectrum envelope of speech with the fine structure of a nonspeech sample recorded by Count Basie in 1939

5) Nonsymbolic

A study using patterned sinewave lures revealed that perceptual organization is nonsymbolic in its effects. That is to say, perceptual organization occurs by sensitivity to speechlike variation, and is distinct from the analytical finesse that creates an impression of linguistic form (Remez, 2001). In this test, the components of a sinewave sentence were arrayed dichotically, separating the tone analog of the first, third and fourth formants in one ear from the analog of the second formant in the other. The challenge to organization was to resolve the coherent variation among the tones composing the sentence despite the spatial dislocation of its constituents. After establishing that listeners tolerate spatial dissimilarity in amalgamating the tones perceptually, a variety of lures was introduced in the frequency band of the second formant in the ear opposite the true tone analog of the sentence, which also contained the tone analogs of the first, third and fourth formants. In this kind of presentation, perceptual organization is challenged to resist the lure presented in the same ear as the first, third and fourth tone, and to appropriate the second tone that completes the sentence. Some lures were easy to resist; those that were constant in frequency or those that alternated brief constant frequency tone segments did not harm the organization of dichotically arrayed components. Other lures were far more difficult to withstand: a gradient of speechlike variation was created by straining (or, more commonly, squashing) the frequency variation of the lure to vary from speechlike to constant frequency at the average frequency of the second formant; the speechlike variant was actually a temporally reversed second formant analog. The lure could not complete the sentence because its variation was never coherent with the first, third and fourth tone. Nor did it evoke impressions of linguistic form; the lure in the clear sounded like a warbling pitch pattern. Nonetheless, it interfered with organization in proportion to its frequency variation, the most when it varied with the pattern of a second formant; less when it varied in frequency at the pace and pattern of a second formant but over a reduced frequency range; and not at all when it was squashed to a constant frequency. In other words, the propensity of the lure to interfere with perceptual organization depended on the speechlike properties of its frequency variation alone, because this single time-varying tone did not evoke phonetic impressions that competed with those of the sentence.

6) Requires attention

A subjective impression of inevitability often accompanies the perception of utterances, but this belies the actuality; the perceptual organization of speech requires attention, whether elicited or exerted. This is seen, again, in studies of sinewave replicas in which the auditory quality is so little like speech listeners are unlikely to organize the tone complexes spontaneously (Remez et al., 1981; Remez, Pardo, Piorkowski, & Rubin, 2001). When a sinewave replica of an utterance is heard simply as a set of contrapuntal tones, none of the auditory qualities is vocal, hence nothing about the experience compels the perception of linguistic properties. Physiological measures were consistent with the hypothesis that no covert aggregation of the tones occurs if the perception of auditory form takes place without apprehension of phonetic attributes (Liebenthal, Binder, Piorkowski, & Remez, 2003). However, a listener who is informed that the tones compose a kind of synthetic speech is readily capable of transcribing a sinewave sentence on this instruction alone; no training, extensive exposure or special hints are needed. In this condition, the aim of hearing the tones phonetically permitted attention to the coherence, albeit abstract, of the speechlike variation in the tones. Moreover, only the sinewave patterns derived from natural utterances are amenable to organization by virtue of the exercise of attention. An arbitrary or incomplete physical spectrum does not evoke an impression of phonetic attributes simply because a perceiver intends to resolve a speech stream. Because natural and much high-quality synthetic speech elicits phonetic attention by virtue of intrinsically vocal auditory qualities, the role of attention in speech is easily overlooked. More generally, this finding of a contingency of organization on attention in the case of speech anticipated the claim that auditory perceptual organization in the main depends on attention (Carlyon, Cusack, Foxton, & Robertson, 2001). It also falsifies the claim that speech perception is accomplished by a modular faculty, because the contingency of perceptual organization on attention contradicts the premise of autonomous and mandatory action immune to influence by belief.

2. Audiovisual Perceptual Organization

The classic formulation of perceptual organization took the cocktail party as the critical setting, and much of the ensuing research examined the proficiency with which a perceiver pulls the speech of a talker of interest from a lively acoustic background. Although this conceptualization has been productive, a different slant is needed to describe a listener who can also see the talker. It has been well established that in this situation a perceiver treats speech as a multimodal event, sampling visually and auditorily (Sumby & Pollack, 1954). In multimodal speech perception, the formal characterization of finding and following a speech stream remains much the same as the auditory instance, with a twist. Rival conceptualizations have characterized multimodal perceptual organization in parallel streams converging at the end or, as a single multimodal stream in which visible and audible features interact continually. To caricature the two perspectives, the first pictures the perceiver as both a blind listener and a deaf viewer huddling within a single skin, resigned to negotiate any discrepancy between utterances that each determines after perception concludes. The second perspective conceives of the perceiver as an auditoryvisual synesthete in whom visible and audible ingredients blend so thoroughly from the start that no residue remains to distinguish the sensory core of the phonetic forms. Whichever conceptualization comes closer to the truth, the visual resolution of speech poses the familiar challenge to organization: the physical effects are regular albeit unpredictable, and none of the effects in detail is unique to speech. This is so whether the level of description is a stream of light reflected from the surfaces of the face, a 2 1/2-d sketch of an as yet unresolved face in a visual scene, or a description of the face as a familiar kind of object in motion.

1) Intersensory combination

In considering the problem of multimodal perception, it is also natural to speculate about the grain of analysis at which intersensory combination occurs (Rosenblum, 2005; Lachs & Pisoni, 2004). Principally, this is a puzzle to solve only if the second alternative conceptualization, of the amalgamated multimodal stream, proves to be true. If speech perception occurs separately in visual and auditory modalities, then perceptual organization proceeds in parallel for visible and audible samples, and any interaction between modalities occurs after perceptual analysis has resolved the phonemic form. Because the phonemic properties are set by their contrastive linguistic function, and not by the specific transient sensory or motor forms of their expression, the dimensions of visually perceived speech intrinsically match those of auditorily perceived speech. Under this condition, the scientific puzzle of understanding the alignment of visible and audible streams is obviated.

The opposite conceptualization, in which intersensory blending occurs in organization, poses a puzzle, for there is no obvious dimension common to vision and hearing. The sensory qualities of these two modalities are largely incommensurate – hue, brightness and saturation do not form tight analogies to pitch, loudness and timbre – and it should be evident that the acoustic transparency permitting a listener to hear the changes originating in the action of the glottis and tongue body have no counterpart in a visible face, in which the larynx and all but the lips and the tip of the tongue are out of sight. Some research proposes the existence of a common intersensory metric, an intermediary permitting the visual and auditory streams to blend in a form exclusive to neither sense (Massaro & Stork, 1998). Variants of this proposal cast auditory sensation – pitch, loudness and timbre – as the common metric into which visual form is also cast for a spoken event (Kuhl, 1991), and a kind of shallow representation of visual and auditory primitives in articulatory parameters (Rosenblum & Gordon, 2001), about which there is much more to say when we turn to phonetic analysis.

Evidence favoring each conceptualization of intersensory relation exists in the technical literature, chiefly in studies of audiovisual merger. That is, synthesis and digital editing of video and audio components have been used to create phonemically discrepant visible and audible presentations, with which to determine the nature of multisensory combination. In the original use of this method (McGurk & McDonald, 1976), an audio [bɑ] and a video [gɑ] were resolved as a fusion, [dɑ]. In tests of this kind, it is possible to fix the identifiability of auditory and visual components independent of tests of their combined effect, and such findings are subsumed well within a parallel model of perceptual organization in which concurrent perceptual states are reconciled if the perceiver organizes them as bound to the same talker. Alternatively, some audiovisual phenomena are simply not well described by parallel and segregated organization in each modality.

In one of these intriguing cases, a video of a face was presented with an amplified electroglottograph signal correlated with the pulsing of the larynx of the depicted talker (Rosen, Fourcin, & Moore, 1981). The appearance of the face was unexceptional; the electroglottograph sounded like an intermittent buzz changing in pitch in the range of the voice. Some of the syllables and words could be resolved phonetically by watching the face, and the audible buzz evoked no impressions of words at all. Overall, the conditions for unimodal visual speech perception were barely met, and were not met at all for unimodal auditory speech perception. In this circumstance, multimodal perception should be poor in as much as the cumulative effect of poor visual and no auditory perception remains poor. Instead, the combination was fine, arguably reflecting the effectiveness of auditory and visual streams in combination when separately neither stream was adequate. This finding among many others offers evidence of a common dimensionality for viewed and heard speech preliminary to analysis.

2) Mismatch tolerance

Among the best clues to the nature of multimodal perceptual organization are the results of studies of the tolerance of spatial and temporal discrepancy across the modalities. In one notion, vision and hearing supply discrepant but complementary samples of speech (Bernstein, Auer, & Moore, 2004), and this simplification describes both audiovisual presentations in which visible and audible patterns coincide and cases of intersensory competition. The organization of fine grain discrepancy between viewed and heard speech scales up to coarse grain discrepancy, and this functional similarity over scale variation is surprising. At the finest grain, auditory and visual streams are mismatched simply because of the disparity in the aspects of the physical acts of articulation that each provides, and not only because the primitives of auditory and visual sensation differ. Indeed, the enterprise of multimodal research rests on findings that fine grain discrepancies introduced by a scientist’s method are resolved in perception, often without eliciting an impression of disparity in the seen and the heard speech. And, at the coarser grain as well as the finer, perceptual integration of discrepant sensory samples is robust.

In several studies, perceptual integration survived spatial displacement of audio and video sufficient to notice (Bertelson, Vroomen, & de Gelder, 1997); and, temporal misalignment of audio and video sufficient to notice (Bertelson et al., 1997; Munhall, Gribble, Sacco, & Ward, 1996). To be more precise, the merging of phonetic features used to index the perceptual integration of vision and hearing persisted under conditions of sizeable spatial and temporal divergence. Such findings can leave the researcher without a convenient explanation because the theory of first resort fails to apply. Specifically, the very tolerance of mismatch blocks the psychologist's automatic and tiresome appeal to similarity as the engine of integration; the integrated streams are dissimilar, displaced and lagged. And, the conditions created within the audiovisual display introduce discrepancies at a scale that surpasses ordinary experience by an order of magnitude. Appeals to likelihood can seem clichéd in psychological explanation, but this procrustean tactic must fail in these instances. The relative divergence of the integrated streams is just unfamiliar.

3) A unimodal and multimodal contour

How, then, does a perceiver apprehend the disparate sensory samples of speech as a coherent progressive event? When organization is veridical, the auditory or visual effects are grouped despite dissimilarity and discontinuity of the sensory constituents of a perceptual stream. The familiar principles of perceptual organization deriving from Gestalt laws of figural organization (Bregman, 1990) cannot be responsible for unimodal organization, for they invoke one or another form of similarity among the sensory constituents. If a role for this conventional account seems unlikely to suffice in unimodal organization, it is utterly implausible for explaining the cases of multimodal organization in which some form of binding appears to occur intermodally in advance of analysis. Although the discussions in the technical literature generally portray binding as a process of sorting analyzed features into bundles coextensive with objects, it appears as though the urgency of auditory perceptual organization compelled by the fast-fading sensory trace imposes a different order. Instead, binding of the sensory constituents of the spoken source must occur before analysis, and perhaps this is the cause of the condition that the perceptual organization of speech requires attention. It is tempting to speculate that there is a single set of organizational functions that applies regardless of the assortment of samples arrayed across the modalities, and some studies of neural metabolism correlated with perceptual organization (Liebenthal et al., 2003) are consistent with this view – but evidence of its existence holds far less value than evidence of its characteristics would.

3. Perceptual Analysis

A perceiver who resolves a stream of speech in a raucous or tranquil scene might also be able to resolve its linguistic form. These facets of perception are contingent. Certainly, the circumstance in which a listener knows that someone is speaking but cannot make out the words is familiar to us, although the inverse – linguistic impressions of a spoken event in the absence of an impression of someone speaking – might merit a thorough reappraisal of mental status. The perceptual resolution of linguistic form has been a topic within the technical study of speech for more than seventy years, and the longevity of this concern is due to the intriguing complexity of this type of sensitivity. Although it has taken a variety of guises, in each the central challenge has been to understand the perceptual ability to apprehend the expression of a small number of linguistic forms under conditions that vary without end.

Long ago, research on the perceptual analysis of speech adopted a focus on the ultimate constituents of language. That is, the linguistic properties that speech expresses are componential, and the components are hierarchically nested. Utterances in the form of sentences are composed of clauses, within which phrases are nested; phrases comprise words, words are composed of syllables and each syllable can be a series of phoneme segments. Phonemes are grouped by distinctive features, that is, by virtue of the coincidence of disjunctive attributes that, together, constitute a system of contrasts across the segmental inventory of a language.

To researchers of the first generation of psycholinguists, the componential nature of linguistic structure was theoretically significant, though the focus on ultimate constituents in speech perception research was also practical (Miller, 1965). It is not sensible to focus on sentences as irreducible objects of perception— there is an infinite number of them, and of phrases, too. Although languages differ in the number of words that they sustain at any moment in history, the set of these is also large. To consider a specific instance, in English, a language often studied in psycholinguistics, words derive from Germanic and from Romance heritages, and for this reason English is said to incorporate more words than is typical of other languages with calmer history. A focus on individuals of a specific group can restrict the study to vocabulary in regular spoken use – in contrast to the far larger recognizable vocabulary – which imposes an inventory of roughly 15,000 items (Miller, 1951). If this is a saving from the infinity of sentences and a large lexical stock, even greater economy is achieved by considering that whatever the word, in English it is composed from a supply of three dozen or so phonemes expressing perhaps a dozen and a half contrast features. Taxonomies of phonemes and the features on which the classes are sorted can become controversial from time to time, depending on the rise or fall in value of one or another kind of evidence. Even with such disputes, there has been good agreement that the perception of speech entails the perceptual resolution of elementary linguistic attributes available in a brief spoken sample; larger structures of linguistic form are produced cognitively by aggregating the elementary constituents provided by speech perception. We defer a discussion right now of the consequences of the phonetic expression of phonemic contrasts, but not for long.

Setting a perceptual focus that is linguistic, segmental and contrastive defines the products of perception, although consensus about the effects has not tempered the disagreements about the causes of perception. This dispute among perspectives concerns the kind of perceptual analysis yielding the linguistic objects. Proponents have divided on its essential nature. Either the perception of speech depends on auditory sensitivity and categorization, or on articulation, or on linguistic function. Each of the proposals is old, and the stalemate is apparently perpetual. We will offer a recommendation, but first we expose some of the technical details.

1) A general auditory account

The roots of the auditory approach run deep. Among the earliest reports in experimental psychology are studies of the likeness of simple whistles and buzzes to speech sounds (Kohler, 1910; Modell & Rich, 1915). Although the correlations were only rough, they licensed the claim that vocality is a primitive auditory sensory quality. The argument held that because vocal impressions are elicited by simple acoustic attributes they are fundamental in human sensory experience; therefore, a talker's ability to evoke a listener's phonemic states depends on producing sounds that hit auditory targets given subjectively and intrinsically. There are more technically sophisticated versions of this antique claim at large today (Kuhl, 1991), but the germ of the idea is similar. Indeed, such findings are perennially welcome in psychology due to a resilient eagerness for sensory reduction of perceptual impressions of the structures and motions of the world. The draw of this explanation is that it permits a description of perception to attribute an incidental role to the objects and events that ultimately cause sensory states: All of the explanatory action pertains to the sensory pathway and neural centers of associative learning. Indeed, it has become commonplace recently for the justification to invoke a perceiver's ability to learn the statistical characteristics of the distribution of sensory states with which phoneme contrasts allegedly coincide. This premise invokes a hypothetical norm in its attempt to accommodate the variability in the acoustic form of each phoneme due to the variety of talkers, rates of speech, and attitudes expressed concurrent to language production, each of which precludes an acoustically uniform expression of a phoneme across different occasions. In one expression of this idea (Diehl, Kluender, Walsh, & Parker, 1991), the auditory system is viewed as a nonlinear conduit of the acoustic effects of speech in which contrast is created by means of enhancement of some auditory elements relative to others. Admittedly, adherence to a general auditory perspective is only weakly justified by psychoacoustics or auditory physiology (Diehl, Lotto, & Holt, 2004).

The perspective on speech perception offered in a general auditory approach has a goal, to pursue a model of the phonemically interested listener as a trainable ear and little else. In a recent review, Diehl et al. (2004) argued that the explanatory detail presently accrued under this rubric is too thin to permit a falsifying test, but this reservation seems unduly gloomy. Even if precise predictions of experimental findings are not readily produced from the principles underlying the approach, it is sensible to ask if the premises of the model attach importance to false assertions. Specifically, if the ambition of the model is not mistaken, its allure is surely diminished by two well established properties of speech perception: (1) the fleeting nature of auditory forms; and, (2) the irrelevance of auditory norms. First, in this class of accounts, perception is based on the varying sensory correlates of speech sounds, and a listener's personal history of experience with /d/, for instance, is encoded to generate a long-term probability distribution in which more and less typical auditory manifestations of /d/ are calibrated. The success of a listener in recognizing instances of this phoneme would necessarily depend on the likelihood that an as yet unidentified sensory form can be assimilated to a longstanding likely auditory representation of /d/ among other segments in the language. But, classic psychoacoustic research revealed that the auditory properties of speech are exceedingly fragile, and are difficult to protect for even a quarter of a second (for instance, Howell & Darwin, 1977). This limit must be a mild embarrassment, at least, to a conceptualization relying on the durability of raw auditory impressions of speech. Although such representations are reasonably chosen for instrumental applications such as speech-to-text systems, these are constrained by circuit design and not by physiology (Klatt, 1989). To survive in a listener's memory, short-lived auditory properties acquire a different form, possibly in in a listener's memory, short-lived auditory properties acquire a different form, possibly in the dynamic dimensions of the sources that produced them (Hirsh, 1988), and when a listener remembers a sound, it is more likely that the recalled quality is generated rather than replayed from a faithful inscription in memory of the original auditory form.

A second problem for a general auditory account of speech perception is its reliance on auditory manifestations of the phoneme contrasts graded by likelihood. Even to entertain this premise, we must be credulous momentarily about the prior claim that unelaborated auditory forms of speech are retained well in memory; this suspension of criticism permits us to review the assertion that a spoken phoneme is identified by a normative assessment of its sensory form. In short, the robustness of intelligibility over widely varying natural conditions of acoustic masking and distortion show clearly that neither goodness nor typicality in auditory quality is requisite for speech perception. Indeed, intelligible sentences are perceived from patterns dissimilar to speech in acoustic detail and in auditory effect (Remez et al., 1981; Shannon et al., 1995; Smith et al., 2002). But, what is the shape of a distribution of the auditory attributes of a phoneme?

To be truthful, no one knows. There is a single study of actual incidence, of the exposure of a single infant to speech produced by one adult (van de Weijer, 1997). This means that claims about sensitivity reciprocating the distributions of the acoustic or auditory forms of speech are hopeful, and without empirical foundation (Saffran, Aslin, & Newport, 1998); at least, no claim is grounded empirically yet. But, it is not difficult to recognize the implausibility of the claim that auditory typicality determines the perception of phonemes. The typical auditory forms of speech must be those sensory states evoked by exposure to the acoustic products of vocalization. After all, an overwhelming majority of instances must be those in which a listener perceives speech because a talker spoke. These days, the pervasiveness of the experience of speech over the telephone also contributes to normative distributions, and so does speech produced by talking toys and gadgets. Overall, the probability distribution must represent this kind of typical experience composed chiefly of acoustic vocal products with minimally distorted variants at the improbable ends of the distribution.

In fact, listeners are evidently not fussy about the acoustic constituents or the auditory qualities of intelligible signals. Neither natural broadband resonances nor harmonic excitation nor aperiodic bursts and frictions nor any specific set of acoustic correlates of a phoneme is required for perception (see Figure 1). Instead, a listener perceives speech as if the commitment to the particular sensory realization of the linguistic contrasts is flexible. This readiness to find functional contrasts in the least expected acoustic or auditory form opposes the fixity of an audit ory norming rationale. Indeed, such acoustic norms – some without auditory warping – form the basis of speech-to-text devices often aimed at the typical expressions of just a single individual (Picheny, 2003); even so, we are still typing.

Before turning to consider an account of perception grounded in articulation, it is useful to note that there are important questions about auditory function in speech that do not depend on the claim that phoneme categories coalesce out of auditory form. At the most elementary, the acoustic correlates of each linguistic contrast are multiple: the speech stream itself is a composite of dissimilar acoustic elements. Attention to the auditory quality of constituents of a speech stream – an aperiodic burst, a second formant frequency transition, a noisy hiss – can occur concurrently with attention to the linguistic properties – an unvoiced fricative of coronal articulatory place. This kind of bistable perception in which attention can hold auditory form or its superordinate, or both, is not well understood outside of musical contexts. Moreover, if qualitative attributes of speech are retained in a durable form, the dimensionality of such knowledge is not well explored (Hirsh, 1988). At the largest grain, the flexibility of the standards for perceiving the linguistic elements of speech is well evident, yet the function by which a perceiver resolves linguistic properties in specific instances, especially those evoking novel auditory form, remains a tough puzzle.

2) An articulation-based account

Modern linguistic description took shape with phoneme contrasts already described in the dimensions of articulation. The technology required to portray acoustic properties did not exist, and in the resort to articulatory dimensions to describe the sounds of speech, Joos (1948) says linguists made a virtue of necessity. However, this practice was unsatisfactory even as articulatory description, largely because the method presumed anatomical and functional states of articulators without direct evidence. For instance, the classical notions of articulatory contrasts in vowel height and advancement were designated by intuition, not by observation, and ultimately proved to be inaccurate portraits of the tongue shape and motion discovered in x-ray fluoroscopy, electromyography and magnetic resonance imaging (Honda, 1996).

When methods for direct measurement of sound became available to supplement impressionistic descriptions, it had a paradoxical effect on the restlessness with old fashioned articulatory description. As the basic properties of speech acoustics were described technically, a problem emerged for proponents of acoustic description; indeed, the conceptualization of articulation was challenged as well. Research on production and perception alike failed to find counterparts to the theoretical description of phonemes in articulatory, acoustic or auditory components. Each perspective in its own way had presupposed that speech was a semaphore, with every phoneme a kind of vocal act or pose, or every segment a kind of acoustic display. Instead, whether construed as acts of articulation, their physical acoustic products or their psychoacoustic effects, an apparent lack of invariance was evident in the correspondence of the linguistically contrastive phoneme segments and their expressive manifestations. In each domain, the relation of a phoneme to its articulatory and acoustic correlates proved to be one-to-many.

Of course, the mere existence of variety among the physical or physiological correlates of a linguistic component is not troublesome to perceptual explanation. If the articulatory, acoustic or auditory tokens of different phonemes correspond uniquely to types, the lack of invariant form is insignificant because the correlates of one type are not shared with any other. The critical finding about the relation of phoneme to correlate was the nonexclusive relation between type and token. One of the clearest instances is the /pi/-/kɑ/-/pu/ phenomenon (Liberman, Delattre, & Cooper, 1952) in which a single acoustic element evokes an impression of a labial consonant and a palatal consonant depending solely on the vowel with which it is presented.

A key explanatory innovation occurred in response to such findings. A new sense of the idea of coarticulation was created to describe the relation of production and perception; a history of coarticulation in phonetic linguistics is offered by Kühnert and Nolan (1999). At the heart of this breakthrough was the inspiration that descriptively segmental phonemes are encoded in articulation (Liberman, Cooper, Shankweiler, & Studdert- Kennedy, 1967). That is to say, a perceived phoneme sequence is restructured rather than produced as a simple sequential cipher or articulation alphabet. The encoding occurs because the vocal articulators are intrinsically separable into controllable parts, and the expression of a sequence of phonemes is thereby reassembled as an imbricated pattern of constituent acts that unfold concurrently and asynchronously. This approach explained well the inexhaustible variety of articulatory and acoustic correlates of each phoneme, or, rather, the lack of a consistent physical manifestation of a phoneme, because whichever segments preceded and followed it shaped its articulation by contributing to the encoding; and, no phoneme is ever expressed in isolation of coarticulatory influence. Liberman et al. explain that such recoding achieves high rates of segmental transmission with sluggish anatomy. The cost is to obscure the relation between an intended or perceived phoneme and its articulatory and acoustic form. Accordingly, speech gives phonetically encoded expression to an intended if abstract phoneme series.

From this premise, a characterization of perception follows as directly as night follows day. If the acoustic speech stream is an encoded version of phonemes, due to the articulatory restructuring of an abstract segmental series, then an inverse operation required to apprehend the segmental series obliges a perceiver to reciprocate the motor encoding in some fashion. A variety of specific technical hypotheses about this kind of perception was ventured in different versions of the motor theory, including reliance on learned articulatory correlates of the auditory forms of speech; covert efferent mimesis; and, imagined surrogates of proprioception that accompanied a talker's speech.

Challenged by evidence, the motor theory looked terrific at a distance, from the perspective of neuropsychology or studies of human evolution (Galantucci, Fowler, & Turvey, in press). At close range, the disconfirming proofs of its technical claims emerged steadily from detailed research on the relation between perception and production. Crucially, studies of extremely young infants showed that perceptual sensitivity develops in advance of articulation, and is not a consequence of it (Jusczyk, 1997). In adults, the invariant characteristics presupposed of the articulation of individual phonemes was falsified in studies of articulatory motion and electromyography (MacNeilage, 1970). This is an enormously intriguing literature impossible to gloss. Yet, acknowledging exceptions, the fair preponderance of evidence showed that every phoneme takes many anatomical forms, and invariance in the correspondence of phoneme to motor expression was found neither in an aggregate of α-efferent activity, nor in the precise motion or configuration of articulators, nor in the shapes of the vocal tract achieved by articulation. In a revision of the motor theory proposed to answer research that it had motivated, perception was held to resolve the invariant phonemic intentions of a talker rather than the acts of articulation, varying without limit, as they are executed (Liberman, & Mattingly, 1985). This version represents spoken communication as a transaction composed of deeply encoded phonemic intentions, aligning the revised motor theory with the symbolic emphasis of a linguistic view of speech perception.

A pair of conjoined accounts of more recent vintage aims to span the gulf between intention and action while retaining the emphasis on production of the motor theory: articulatory phonology and direct realism (Goldstein & Fowler, 2003). Articulatory phonology offers a description of linguistic contrast set in abstract articulatory primitives, and direct realism describes a perceiver's sensitivity to articulatory contrasts by attention to the visible, audible and palpable effects of speech. The crucial contribution of this proposal is a representation of lexical contrasts in a repertoire of gestures, not of phonemic segments. Building on the characterization of articulation given by Liberman et al. (1967), a contrastive gesture is designated as: (1) a movement of a particular set of articulators; (2) toward a location in the vocal tract where a constriction occurs; (3) with a specific degree of constriction; and, (4) occurring in a characteristic dynamic manner. In this perspective, a word is indexed by a gestural score describing its production as the coupled asynchronous action of lips, tongue tip, tongue body, velum and larynx. Such gestural components are understood as quasi-independent actions of vocal articulators. The pattern with which gestures impose and release constrictions creates the contrasts customarily described in a segmental phoneme series. The traditional separation of phonemic contrast and phonetic expression theoretically collapses in this account into an equivalence between linguistic properties and vocal acts. With respect to the principle at the core of the motor theory, this asserted equivalence of linguistic contrast and manifest articulation denies the encoding that supplied the articulatory character of the inverse function purportedly applied by a perceiver to a speech stream. In complementary function, the account describes the perceiver distinguishing words in the same gestural components that the talker employs to create speech. The phonemic properties of speech are apprehended perceptually without decoding them, according to this argument, because the acoustic pattern and its sensory effects are transparent to the articulatory components that index spoken words across the lexicon.

A gestural score representing the coupled actions of quasi-independent vocal articulators in the production of the word SPAM. (Browman & Goldstein, 1991, p. 318). canonical phoneme sequence actually varies in exact phonetic, or expressed, detail. The word SECURITY, for example, is produced as these variants, among others: [səkhjʊɹithi], [skhjʊɹəɾi] and [skhjɾi]. Under an articulatory phonology, many variants are potentially rationalized as consequences of minimally different task dynamics of the same gestural components given in a lexical representation. Variation attributable to differences in speech-rate, reductions, lenitions and apparent deletions are likewise described as natural variants of the same gestural form, and this sameness remains available to perception, in principle. A listener who resolves the gestural components in a speech stream might notice but pay little attention to the effects of slight phase differences in the expression of constituents of an intended contrast. Even in casual speech, in which the canonical forms of words can be compromised by the expressive aims of a talker, articulatory phonology promises to explain the variation without representing the phonetic form of the articulation differently from the phonemic form distinguishing a word from all others. Admittedly, in some synchronic and diachronic cases, it seems that talkers do express different contrasts than the lexicon employs, yet the different representations warranted by these facts can nonetheless be described by postulating no more than minimal changes in the components of a gestural score. Of course, there are some phonological phenomena in some languages that defy simple characterization – Nature provides her own exceptions – and it is not clear how these will be resolvable to general principle in relating the phonemic and expressed forms (Browman & Goldstein, 1991). But, the perceptual claims of this account are readily evaluated.

Two critical axioms are assumed in the perceptual account given by articulatory phonology and direct realism, and if they are not exactly false, they are less true than the account demands. The first is an asserted isomorphism between the components used in language to create contrast and the components of spoken acts; they are designated as a single set of gestures. The second is a state of parity such that talker and listener match; the expressed forms and the perceived forms are claimed to be the same. Of course, these axioms are related. People who speak the same language use the same canonical contrasts. If they express them differently, or if a talker nonaccidentally expresses the same form in gestural variants on different occasions, then the relation between canonical and expressed forms can be regulated, adjusted or reshaped; phonemic and phonetic form are not identical in this circumstance. Instead, some of the degrees of freedom in articulation would be reserved for expression beyond those that are committed to the canonical form of the word. If articulation varies with a talker's communicative aims, then canonical and expressed forms do not match, and parity must be achieved, not simply fulfilled.

The axiom of parity denotes sameness in language forms shared by a talker in composing an expression and by a listener in perceiving it. Here, the intended sense of parity applies only to the gestural components of language, and this is completely apt. After all, lexical parity is commonly breached: conversations are fine with people who say LIGHTNING BUG when we say FIREFLY, SACK for BAG, TRASH for GARBAGE, that is, under conditions in which the communicative function matches while the lexical form does not. The claim of parity states that the forms of perception and production are the same, and the claim at some level of resolution cannot be false (Liberman, 1996). If the perceiver knows what the talker said well enough to repeat it, lexical and phonemic parity occurred, at least. But, because isomorphism is suspicious, expressed forms can be understood to differ from abstract phonemic forms. The axiom becomes harder to sustain in that case because phonemic parity can occur without phonetic parity; and, because phonetic parity is so unlikely, even in monozygotic twins reared in the same household (Johnson & Azara, 2000; Nolan & Oh, 1996; cf. Gedda, Bianchi, & Bianchi-Neroni, 1955).

Critical data on this topic indicate that perception is bistable, permitting attention to be drawn to superficial and canonical form concurrently. The study (Goldinger, 1998) used an original measure of perceptual resolution. An experimenter presented a recorded utterance for a subject to repeat immediately or after a brief imposed delay. Comparing the elicited speech samples to the eliciting sample showed that utterances produced immediately were more similar to the eliciting sample than were those produced after even a brief delay. Despite all, a similar utterance was far from a faithful replica of the model. This is expected, to be precise, for not even nightclub impressionists achieve their characterizations by exact replication of the speech of John Wayne and Cary Grant – and they rehearse. The difference in the two conditions of lag must be attributed to phonetic attributes inasmuch as the words were the same, hence, the contrastive phonemic properties were the same. With respect to the parity axiom, though, the result is troubling, because the finding of only rough similarity insinuates that if parity is fostered it is unattained; and, that the faint shadow of parity that actually is manifest lasts only a moment, and once the impulse toward parity subsides the default state of disparity returns. Moreover, studies of deliberate imitation show that an individual typically provides an erroneous imitation of a self-produced speech sample (Vallabha & Tuller, 2004). If parity does not occur in this limiting case, when would it?

A plausible description of these phenomena is possible without invoking a disposition to isomorphism and parity. In the moment when a spoken word is perceived, phonemic and phonetic forms are resolvably different from each other. The salient differences often include aspects of a talker's speech that differ greatly or minimally from a perceiver's own characteristic articulation. Speech initiated in this state can be nudged toward the form of the immediate phonetic model and away from the habitual expression of canonical form. At a greater delay, though, the vividness of the phonetic impression of the eliciting utterance has faded in its contrast with long established articulatory habits, and production is free from the adulterating pressure of a phonetic form distinct from the talker's intrinsic dynamic. It is as if talker and listener express lexical and phonemic parity by means of their phonetic differences. With sustained exposure to an individual talker, a perceiver is likely to form an impression of the talker's characteristic articulatory variation, sufficient to imagine speech produced in the voice and style of the talker, and perhaps to adopt the phonetic characteristics in a deliberate imitation (Johnson, Foley, & Leach, 1988). There is some evidence that such vicarious experience of the speech of familiar others can influence a talker's production in detail (Sancier & Fowler, 1997). But, the listener and talker need not match phonetically for any of this to occur. Indeed, in order for the phonetic similarity of two talkers to wax and wane, they cannot match.

The assertions of isomorphism and parity mask a significant aspect of the perception and the production of speech, namely, the ubiquity of mismatching form. Whether the discrepancy occurs in the visible and audible properties of speech, as in audiovisual speech perception, or in the phonetic realization of phoneme contrasts, as occurs whenever two individuals speak to each other, it seems that you neither expect nor require your conversational partners to use the identical expressive forms that you use. Or, more precisely, the sharing of words apparently licenses variegation in articulation, both in groups – as dialect unless the group also possesses an army and a navy, in which case it is a language – and in individuals – as idiolect.

3) Perceiving speech linguistically

A linguistic emphasis in explanations of speech perception is familiar. The basic notion deriving from Jakobson and Halle (1956) identifies phoneme contrasts as symbolic and linguistic, and neither articulatory nor auditory. In this regard, they assert symbolic status to the phoneme and the word alike. This is subtle, for it warrants a distinction between the form of words ("I said PIN, not PEN") and their meanings ("I meant PIN, not PEN"). The relation between sound and meaning is arbitrary notwithstanding the contrary claim of phonesthesia, a perennial topic of romantic symbolists (Aman, 1980). In order for the listener to know what the talker meant, the listener must resolve the form of the talker's utterance; without grasping the form of a talker's speech, a listener has merely guessed the talker's meaning. It is this juncture that is critical for this conceptualization, because of the complexity in the relation between the canonical form regulated by the language and the expressed form regulated in compromise between linguistic and personal expression.

Initially, accounts of this genre offered a well-defended description of perception as a process of increasing abstraction (cf. Halle, 1985). The difference between phonetic form and canonical phonemic form decreed the initial conditions. Perception began with a sensory pattern, and the perceiver was obliged to transform it in order to resolve its phonemic attributes. The asynchronous distribution of acoustic correlates of a phoneme in a speech stream precludes a simple alignment of the sensory attributes and a canonical segmental series. In this model, several influences on the expressed form of speech must be undone before the segments can be exposed: the effects on the acoustic correlates of phoneme contrasts due to variation in the rate of production, the effects attributable to anatomical scale differences among talkers, the effects due to differential placement of emphasis, to variation in articulatory clarity, to foreign accent, and, of course, the effects due to co-production of sequential phonemes, syllables and words. In short, the characterization depicted a perceiver wielding stable standards – schemas – of the typical sensory presentation of the phonemes in the language, and applying a perceptual function to strip the instance-specific detail from an impinging sensory stream. Once a sensory sample was recast with sufficient abstractness, it was fit to match a stable linguistically-determined form.

Evidence from the listening lab had calibrated a perceiver's suppleness in adapting to the properties that drive the expressed form of speech to depart from a hypothetical abstract form. If some proposals relied on a dynamic that operated feature by feature (Stevens, 1990), others described the comparison of segmental instances to prototypes (Samuel, 1982), and, in contrast to principles of likelihood, other accounts invoked a standard of segmental goodness independent of typicality yet still subject to the influence of experience (Iverson & Kuhl, 1995). The shared premise of these accounts is the use of progressive abstraction for the perceptual accommodation to variability. Categories of phonemic experience are rightly understood as commutable markers of contrast independent of talker or circumstance. After all, there is no pair of words that depends for its contrast on production by a specific talker, at a specific speech rate, paralinguistic expression of affect, vocal pitch, etc. But, the actual phonetic form of speech is too bound to the local conditions of production to be simply redeemed as an abstract phoneme series composing a word. The view that the incommensurate phonetic and phonemic forms are harmonized by reshaping the phonetic form into a less specific and more general version has been called abstractionist (Richardson-Klavehn & Bjork, 1988).

In such accounts, to appraise an unanalyzed bit of speech a perceiver must reconstruct the incoming sensory form to permit contact with a schematic idealization, or so an abstracting account could claim until critical studies of priming with spoken words. In a priming paradigm, the effect of a collateral probe (called "the prime") on the performance of a perceptual task is generally taken as evidence of relatedness. The closer the relation of a prime and a target, the greater the facilitation by the prime of a test subject's act concerning the target. This description of perception as the recognition of an abstract form warranted equivalence of the detailed phonetic variants of a spoken word used as prime and target because the point of contact inherent to identification was allegedly indifferent to the disparity among spoken instances of the same canonical phonemic form. But, in a series of studies that dislodged abstraction as the orthodox formula in speech perception, test subjects proved to be acutely sensitive to the exact phonetic similarity of prime and target, as if the specific phonetic attributes were preserved, and not simply registered as a preliminary to the process of abstraction requisite to identification (Goldinger, Luce, Pisoni, & Marcario, 1992; Luce, Goldinger, Auer, & Vitevitch, 2000).

In a description of perception by abstraction, the set of contact points is given by the number of resolvable types. The set is potentially small if the segmental phoneme inventory is used. If legal pairs or triads of phonetic segments are used, the set is larger, perhaps tens of thousands for English in comparison to the three dozen phoneme segments, but this set size can hardly be taxing on a nervous system capable of impressive feats of rote learning. But, imagine an indexing scheme representing instance-specific variation: it expands without limit. In contrast to the notion of the infinite use of finite means at the heart of every generative system, long-term knowledge that only encoded every raw instance is simply not compatible with the componential nature of phonology and morphology not to mention parity at any level. And, this consideration cannot apply solely to perceived form, for some studies had shown that we track the differential likelihood associated with the modality of the instances (Gaygen & Luce, 1998). That is, a spoken instance is encoded in a form distinct from a heard instance; a typed instance is marked in memory to distinguish it from a read instance. So far, there has been general agreement that these varied instances coalesce into types that match the abstract forms, preserving the linguistic drivers of differentiation of lexical items through highly varied realization of canonical form. But, how are instances encoded?

In some descriptions of the adaptive resolution of superordinate phonemic types and subordinate phonetic instances, each level is treated as a linguistic representation derived from a raw sensory sample (Goldinger & Azuma, 2003). The instance is preserved as an unelaborated residue of stimulation. A literal understanding of an instance specific memory of utterances warrants a sensory encoding, for this is the only kind of representation that does not oblige the perceiver to an interpretation that substitutes for the direct experience of the instance. Yet, this notion can only be sustained in disregard of the psychoacoustic benchmarks of speech sounds (for instance, Pisoni & Tash, 1974). The unelaborated impression is gone in a tick of the clock. Indeed, the fleeting trace of an utterance arguably forces the retention of instance specific attributes while precluding an encoding of raw auditory experience.

We do not know the form of instance specific attributes yet, though some studies show that a perceiver is exquisitely sensitive to subtle phonetic variants, those that are far more detailed than simple categorization requires (McLennan, Luce, & Charles-Luce, 2003). Some phonetic variants are obviously due to chance–speech produced with food in the mouth, for instance, includes concurrent acts that compromise the expression of linguistic and paralinguistic properties with the accidental moment by moment acts to retain the bolus of food in the mouth. Other subtle phonetic variants are regulated, such as those that distinguish dialects and idiolects, and from their consistency we can infer that their production is perceptually monitored, and that phonetic perception incorporates dialectal and idiolectal dispositions, at least some of the time. Such sensitivity to varieties of phonetic expression at large in a language community might have played a role in findings that subphonemic discrimination of speech sounds always exceeded a prediction based on phoneme identification (see Liberman, 1957). Although these reports had been explained as an expression of auditory sensitivity, fine grain phonetic differences exist at a parallel level of resolution, and it is likely that perceivers attend to this detail because at this grain the linguistic and paralinguistic drivers of expression converge. A finding of instance specificity is potentially reducible to allophonic specificity, at least in linguistic dimensions of this phenomenon. But, not all specificity will be reducible to linguistically regulated properties of speech. In order to explain episodic properties of utterances – you were standing in the moonlight, the breeze was lightly rustling the leaves and a firefly twinkled just as you whispered, "Jazz and swing fans like fast music" – a state-dependent form of inscription might be required, but this is unlikely to be central to language. If this approach to speech perception holds potential for explaining the core pro problems that motivate research, perhaps because it is the most freewheeling of the accounts we have considered. Others lack the suppleness required by the accumulated evidence of the perception of speech as a cognitive function that finds linguistically specified contrasts under conditions that defy simple acoustic, articulatory, visual and tactile designation. A listener who attends to subtle varieties of phonetic expression in speech is obliged to do so by the lack of uniformity in speech production. In accommodating this aspect of variation, a listener meets a challenge created by language communities. The individuals who compose our communities vary in anatomical scale, dialect and idiolect, age, social role and attitude, and these dimensions are expressed in each utterance along with the linguistic message. If the sensory samples reflect these converging influences on expression, it is not surprising that a listener's attention to the attributes of a spoken event include features of the talker and the conditions in which an utterance occurred. In perceiving speech, a listener attends to personal attributes of the talker, and research on the perception of individuals, though it has sometimes run parallel to studies of linguistic perception, ultimately converges with it.

B. Speech Perception In Childhood

1. The Gap

There is a rather glaring gap in our knowledge about the development of speech perception. Whereas a great deal has been learned in the last 30 years regarding infant perception (since the seminal study of Eimas et al., 1971, in which young infants' categorical discrimination for an English phonemic contrast was demonstrated), much less is known about perception in the 16 years or so that intervene between infancy and adulthood. Developmental researchers' engrossment with infancy can be attributed to theoretical, methodological, and empirical factors, including the opportunity afforded to determine those abilities given by nature, the implementation of increasingly sophisticated testing procedures, and the positive findings that have obtained (Bornstein, 1992). To this we may add the recent rise of developmental neuroscience and claims about the special, even overriding importance of development within the first three years of life (cf. Bruer, 1999; Kuhl, 2000).There is a rather glaring gap in our knowledge about the development of speech perception. Whereas a great deal has been learned in the last 30 years regarding infant perception (since the seminal study of Eimas et al., 1971, in which young infants' categorical discrimination for an English phonemic contrast was demonstrated), much less is known about perception in the 16 years or so that intervene between infancy and adulthood. Developmental researchers' engrossment with infancy can be attributed to theoretical, methodological, and empirical factors, including the opportunity afforded to determine those abilities given by nature, the implementation of increasingly sophisticated testing procedures, and the positive findings that have obtained (Bornstein, 1992). To this we may add the recent rise of developmental neuroscience and claims about the special, even overriding importance of development within the first three years of life (cf. Bruer, 1999; Kuhl, 2000).

What we have learned about infant speech perception is indeed impressive (for reviews, see Aslin, Jusczyk, & Pisoni, 1998; Jusczyk, 1997). One fairly wellestablished finding is that early development entails a shift from a languagegeneral to a language-specific pattern of perception (cf. Nittrouer, 2001; Polka, Colantonio, & Sundara, 2001). That is, infants are sensitive at the outset to a wide variety of phonological structures and so are prepared to learn any language to which they might be habitually exposed, but then sometime over the first year of life, sensitivity to many non-native sounds declines. For example, Werker and Tees (1984) showed that 6- to 8-month-olds from English-speaking homes were able to discriminate Hindi consonantal contrasts, as well as those in a Native Canadian language (Nthlakapmx), but that by 9–10 months, this sensitivity had begun to wane – and that by 11–12 months, these same infants no longer attended to these distinctions. (In contrast, older infants from these other language backgrounds could still discriminate these contrasts.) This sort of developmental loss or pruning was attributed to the advent of contrastive phonology around 9–12 months of age, when the infant begins to focus attention on those sounds in the native language that are crucial for distinguishing differences in word meaning (see also Jusczyk, 1993; Stager & Werker, 2000).

However, Kuhl et al. (1992) subsequently found that infants exhibit a "perceptual magnet" effect by 6 months of age (for a full description and critical review, see Walley & Sloane, 2001). Specifically, American infants equate (or fail to discriminate) an English vowel prototype /i/ and its variants, whereas they display better discrimination for a Swedish vowel prototype /y/ and its variants; conversely, Swedish infants equate the Swedish stimuli, but not the English ones. Thus, native language influences for vowels are evident well before 9–12 months, or the point at which it is generally thought that speech sounds first become interfaced with meaning, and infants thus gain entry to the native language proper. Kuhl and colleagues therefore maintained that the infant's initial attunement to the segmental properties of the native language occurs independently of early word learning and is the result of simple exposure to the distributional properties of sounds. Both of these theoretical stances have been challenged (to varying extents) by recent evidence about when, more precisely, infants begin to link sound and meaning.

Unfortunately, the overall impression that is left by much of this research is that little, if anything, of import happens in terms of perceptual development beyond 1 or 2 years of age. Yet one major theoretical reason to expect changes has been alluded to already – namely, during early and middle childhood, there is substantial vocabulary growth (e.g., increases in the size of the lexicon or the number of words that are known) (Anglin, 1993). Such growth in the child's lexical or knowledge base should necessitate changes (perhaps of a qualitative nature) in the way that speech patterns are represented and/or processed. A second reason to expect perceptual advance in childhood is that there is continued exposure to the native language, which might have more subtle (perhaps quantitative), but nevertheless important influences. This expectation is consistent with second-language learning research which has indicated that the phonological system is quite open or flexible up until about 7 years of age (see Walley & Flege, 1999). Third, the reading task with which young children in literate cultures are confronted (especially those who must master an alphabetic writing system) might be expected to have a significant impact on phonological representations and/or processing (see Goswami, 2000).

Little attention has, however, been directed toward speech perception in typically-developing children, so that this would seem to constitute the weakest link in our understanding of the growth of speech perception. In the following sections, I will selectively review what is known about speech perception in childhood. I will briefly outline a model of the development of spoken word recognition (the Lexical Restructuring Model; see Metsala & Walley, 1998) and then highlight the extent to which it is supported by existing data. This model focuses on the impact of spoken vocabulary growth in early and middle childhood, and also provides a framework for understanding changing interactions between phonetic and lexical levels of processing, as well as phonological awareness and early reading success. Finally, I will note some places where the model appears to fall short and identify other important gaps in what we know about speech perception in childhood.

2. Filling the Gap: The Lexical Restructuring Model

Largely because of the influence of infant perception studies (from Eimas et al., 1971, through Kuhl et al., 1992), there has been a substantial, ingrained theoretical bias among developmental researchers that phonetic/phonemic segments are present and functional as units of perception from early infancy. For example, according to Native Language Magnet theory (Kuhl, 1993, p. 133), "[in part because] infants exhibit a language-specific magnet effect at the level of phonetic segments," their speech representations must be "sufficiently finegrained to allow segments to be individuated." Yet this claim about segmental perception has not been directly evaluated in most studies, including demonstrations of the perceptual magnet effect. That is, studies of the effect have typically involved the presentation of isolated vowel stimuli, and so it is unclear whether infants represent/process these stimuli as segments per se or as whole syllables.

Over the last decade, there has, in fact, been a growing consensus that infants' speech representations are not, at the outset, organized around individual phonetic/phonemic segments. Instead, these representations are initially holistic (i.e., based on larger units, such as the syllable) and only gradually, in early through middle childhood, do they become more fully specified and/or undergo segmental restructuring (for review, see Metsala & Walley, 1998; Walley, 1993b). Our Lexical Restructuring Model (LRM; Metsala & Walley, 1998) emphasizes the role of vocabulary growth in prompting such changes for the representation and/or processing of spoken words. In the model, vocabulary growth includes increases in the overall size of the mental lexicon (or the number of words that are known), as well as changes in the familiarity and phonological similarity relations of individual lexical items. Specific expectations about the impact of these factors on spoken word recognition and relevant empirical evidence will be considered below.

In addition, LRM seeks to explicate the relations between children's spoken word recognition, phonological awareness, and beginning reading ability. According to the model, phonemic segments develop gradually as implicit perceptual units for basic speech perception and spoken word recognition, and only later as explicit cognitive units that can be harnessed for the reading task (see also Fowler, 1991; Stanovich, 1988). By this emergent view, phonological awareness, especially the ability to access and manipulate phonemes, is not simply a problem of recovering existing units of speech representation, as the traditional accessibility position has maintained (e.g., Liberman, Shankweiler, & Liberman, 1989; Rozin & Gleitman, 1977); rather, such awareness, which is crucial for learning letter-sound rules, is initially limited by the very nature or developmental status of underlying speech representations (for more complete references, see Garlock, Walley, &

Metsala, 2001; Metsala & Walley, 1998). Some of the more relevant data bearing on the emergent position will be discussed below.

3. The development of phonetic perception

Infants' discrimination of various phonetic/phonemic contrasts does not necessarily involve the detection of localized, segmental differences; rather, discrimination might be mediated by more holistic processes. One of the best demonstrations to this effect can be found in the work of Jusczyk and colleagues (see Jusczyk, 1993), who showed that when 2-month-old infants are familiarized with a stimulus set, such as /bi ba bo/, they are equally likely to dishabituate to /du/ and /bu/ – i.e., they treat /bu/ as novel, even though it shares a consonantal segment with the habituation stimuli. In contrast, young infants do seem to extract or retain some memory for a shared syllable. Sometime between 6 and 9 months of age, infants are, as we have seen, becoming attuned to the consonants and vowels of their native language. Around 9 months, they also begin to display sensitivity to subsyllabic information, such as shared initial consonants and consonant-vowel combinations (Jusczyk, Goodman, & Baumann, 1999), and to the phonotactic patterns, or sequential arrangement of phonetic segments, of their native language (Jusczyk, Luce, & Charles-Luce, 1994). However, the fact that such early sensitivity is evident under optimal testing conditions (e.g., given repeated presentations of stimuli in the clear) does not necessarily mean that the abilities revealed by these tests are robust ones or that development is complete.

In fact, a number of studies point to extant developmental differences in the perception of both vowels and consonants. For example, 3-year-olds' perception of synthetic vowels (/æ/ and /√/) is more dependent on dynamic spectral change information than adults' (Murphy, Shea, & Aslin, 1989) and 5- to 11-year-olds' perception of /i/, /a/ and /u/ is more influenced by stimulus duration, as well as consonantal context (Ohde, Haley, & McMahon, 1996). The highly context-dependent nature of children's perception is better documented for consonants (see Walley & Flege, 1999). In a study by Nittrouer and StuddertKennedy (1987), 3- to 5-year-olds' identifications of syllable initial fricatives from a synthetic /s-S/ continuum were more influenced by vocalic transitions than 7-year-olds' and adults'; these older subjects were more sensitive to the frequency information in the fricative noise. Similarly, Krause (1982) found that 3-year-olds needed a larger difference than adults in preceding vowel length to identify stimuli ending in voiced and voiceless stops, and other researchers have observed that young children pay particular attention to formant transitions in judging place of stop consonant articulation (e.g., Ohde et al., 1995; Walley & Carrell, 1983).

A corollary finding in these and other past studies is that consonant perception by children up to about 5 or 6 years of age appears less categorical than perception by adults; specifically, the slopes of children's identification functions for various stimulus continua are shallower than adults' (e.g., Burnham, Earnshaw, & Clark, 1991; for additional references, see Hazan & Barrett, 2000; Walley & Flege, 1999). More recent investigations have indicated that there are further increases in the consistency of vowel and consonant categorization, including a steepening of the slopes of identification functions, into late childhood and even early adolescence, as well as gains in the ability to make use of impoverished acoustic-phonetic information (e.g., Hazan & Barrett, 2000; Johnson, 2000). Still other studies have shown that young children (about 5 years of age) classify speech patterns on the basis of overall similarity relations, whereas older listeners use phoneme identity (e.g., Treiman & Breaux, 1982; Walley, Smith, & Jusczyk, 1986; see also Section 18.2.3).

Together, these studies suggest that children's representations for speech patterns are not yet adult-like; i.e., their representations are not as fine-grained or segmental, but are instead more holistic in nature and based to a greater extent on information distributed throughout the speech waveform. According to the Developmental Weighting Shift model (see Nittrouer et al., 2000), young children rely to a greater extent than adults on dynamic cues in making phonetic decisions because they are more focused on the recovery of syllabic structure; only with maturation and additional linguistic experience does their weighting of various acoustic properties come to resemble that of adults more closely and be more flexible. This shift is seen to be precipitated, in part, by lexical growth (see also Fowler, 1991); thus, there seems to be fairly widespread agreement that as the lexicon grows, greater attention to the details of the speech signal are required. Despite the transparent nature of this claim, definitive empirical support is lacking. In particular, there have been few developmental studies of how phonetic perception is influenced by lexical status and/or word familiarity.

One exception is a study by Walley and Flege (1999), in which American English 5-year-olds, 9-year-olds, and adults identified synthetic stimuli on a native vowel continuum ranging from /I/ to /i/ and a foreign continuum ranging from /I/ to a foreign vowel /Y/ (presented in a nonword, /C_C/ context). No marked age differences in the location of phoneme boundaries were found – a result that is consistent with the work of Kuhl and others (e.g., Kuhl et al., 1992), suggesting that the vowel space is partitioned quite early in infancy vis-à-vis the native language. However, the slopes of subjects' identification functions became progressively steeper with age, especially for the native continuum. This result is consistent with the notion that young children's perception is not as fine-grained or segmental as that of older listeners. Yet when the stimuli were presented in the context of highly familiar words (i.e., "beep" and "bib"), then young children's slopes were much more similar to those of older listeners. Thus, developmental differences in how sharply defined phonemic category boundaries are may depend, in part, on variations in lexical knowledge. More generally, there are potentially important perceptual-cognitive/linguistic interactions in childhood that remain to be examined.

C. Categorical Perception

Categorical perception was an early finding in the history of the study of speech perception by experimental psychologists (Liberman et al., 1957). When listeners were asked to identify members of an acoustic continuum of syllables varying in the F2 transition that ranged from /be/ to /de/ to /ge/, instead of showing a gradual shift in responses, they showed abrupt shifts, shown schematically in Figure 1.2. This occurred despite the fact that there was an equivalent acoustic change at every step along the continuum. A second hallmark of categorical perception, also shown in Figure 1.2, is that discrimination was considerably worse for pairs of syllables labeled as the same syllable than for syllables labeled differently. An early interpretation of this pair of findings was that it indexed a special way of perceiving speech. According to the motor theory of speech perception, listeners do not perceive the acoustic signal, but rather the articulatory gestures that produced the signal. Categorically distinct vocal tract gestures produce /b/, /d/, and /g/. Accordingly, they are perceived categorically as well. Identification functions are sharp, by this early account, because continuum members with the lowest frequency second formant onsets are perceived as bilabial (on the left side of Figure 1.2). Eventually, a syllable is encountered that cannot have been produced by lip closure, and it and the next few syllables are perceived as alveolar; final syllables all must have been produced by the tongue body, and are perceived as velar. Discrimination is near chance within these categories, according to the account, because all category members are perceived as equally bilabial (or alveolar or velar). It is only when one stimulus, say, is perceived as bilabial and one as alveolar that discrimination is possible. The categorical nature of speech perception has also been challenged by the findings considered next.

D. Duplex Perception

When all of a syllable that is ambiguous between /da/ and /ga/ is presented to the left ear, and the disambiguating third formant transition is presented to the right ear, listeners hear two things at once (e.g., Mann and Liberman, 1983). They hear /da/ or /ga/ depending on which third formant transition has been presented, and they hear the transition as such, as a chirp that either rises or falls in pitch and that is distinct from the phonetic percept. Mann and Liberman interpreted this as showing that there are two auditory perceptual systems. Otherwise how could the same third formant transition be heard in two ways at the same time? One perceptual system renders a phonetic percept of /d/ or /g/. The other hears the transition literally as a fall or rise in pitch. This interpretation has been challenged, but not entirely successfully, by showing that perception of slamming doors can meet most, but not all, criteria for duplexity (Fowler and Rosenblum, 1990). If slamming door parts can be perceived in two ways at the same time, it cannot be because two perceptual systems, a door-perceiving system and the auditory system, underlie the percepts.

E. Multimodal Speech

As noted previously, perceivers of speech use more than acoustic information if they encounter it. They perceive speech in noise better if they can see the face of a speaker than if they cannot (Sumby and Pollack, 1954). Moreover, they integrate acoustic and optical speech information in the McGurk effect (McGurk and MacDonald, 1976). If haptic information replaces optical information, a McGurk effect also occurs (Fowler and Dekle, 1991). These effects may occur because listeners hear and see speech gestures. They integrate across the modalities, because the gestures specified by the two modalities of information should be from the same speech event. Alternatively (excepting perhaps the findings of Fowler and Dekle, 1991), the effects may occur because listeners/ observers have a lifetime of experience both seeing and hearing speech, so they know what it looks like when various acoustic speech signals are produced. Seeing the speaking face, then, helps them to identify what was said. Remarkably, Munhall and colleagues (Munhall et al., 2004) have shown that perceivers can extract phonetic information from the visible head movements of a speaker, such that speech is more intelligible in noise when natural head movements as well as facial phonetic gestures are visible to a speaker. Perceivers of speech are information omnivores. This finding awaits interpretation from either a gestural or an auditory theoretical perspective.

F. Speech Perception And Lexical Access

Psycholinguists tend to classify speech perception and spoken word recognition (and other levels of description) as distinct aspects of spoken language understanding, with most theories postulating distinct levels of representation for each division. Even in theories that posit distinct phonetic-perceptual and lexical levels, the interface between these levels is of great importance. We will focus on three key issues: the lexical segmentation problem, interface representations, and the modularity/interaction debate.

CONCLUSION

Theories of speech perception have often conceptualized the earliest stages of auditory processing of speech to be independent of higher level linguistic and cognitive processing. In many respects this kind of approach (e.g., in Shortlist B) treats the phonetic processing of auditory inputs as a passive system in which acoustic patterns are directly mapped onto phonetic features or categories, albeit with some distribution of performance. Such theories treat the distributions of input phonetic properties as relatively immutable. However, our argument is that even early auditory processes are subject to descending attentional control and active processing. Just as echolocation in the bat is explained by a cortofugal system in which cortical and subcortical structures are viewed as processing cotemporaneously and interactively, the idea that descending projects from cortex to thalamus and to the cochlea provide a neural substrate for cortical tuning of auditory inputs. Descending projections from the lateral olivary complex to the inner hair cells and from the medial olivary complex to the outer hair cells provide a potential basis for changing auditory encoding in real time as a result of shifts of attention. This kind of mechanism could support the kinds of effects seen in increased auditory brainstem response fidelity to acoustic input following training.

Understanding speech perception as an active process suggests that learning or plasticity is not simply a higher-level process grafted on top of word recognition. Rather the kinds of mechanisms involved in shifting attention to relevant acoustic cues for phoneme perception are needed for tuning speech perception to the specific vocal characteristics of a new speaker or to cope with distortion of speech or noise in the environment. Given that such plasticity is linked to attention and working memory, we argue that speech perception is inherently a cognitive process, even in terms of the involvement of sensory encoding. This has implications for remediation of hearing loss either with augmentative aids or therapy. First, understanding the cognitive abilities (e.g., working memory capacity, attention control etc.) may provide guidance on how to design a training program by providing different kinds of sensory cues that are correlated or reducing the cognitive demands of training. Second, increasing sensory variability within the limits of individual tolerance should be part of a therapeutic program. Third, understanding the sleep practice of participants using sleep logs, record of drug and alcohol consumption, and exercise are important to the consolidation of learning. If speech perception is continuously plastic but there are limitations based on prior experiences and cognitive capacities, this shapes the basic nature of remediation of hearing loss in a number of different ways.

Finally, we would note that there is a dissociation among the three classes of models that are relevant to understanding speech perception as an active process. Although cognitive models of spoken word processing (have been developed to include some plasticity and to account for different patterns of the influence of lexical knowledge, even the most recent versions do not specifically account for active processing of auditory input. It is true that some models have attempted to account for active processing below the level of phonemes these models not been related or compared systematically to the kinds of models emerging from neuroscience research.

First, while the cognitive models mention learning and even model it, and the neural models refer to some aspects of learning, these models do not relate to the two-process learning models. Although CLS focuses on episodic memory and focus on category learning, two process models involving either hippocampus, basal ganglia, or cerebellum as a fast associator and cortico-cortical connections as a slower more robust learning system, have garnered substantial interest and research support. Yet learning in the models of speech recognition has yet to seriously address the neural bases of learning and memory except descriptively. This points to a second important omission. All of the speech recognition models are cortical models. There is no serious consideration to the role of the thalamus, amygdala, hippocampus, cerebellum or other structures in these models. In taking a corticocentric view these models exhibit an unrealistic myopia about neural explanations of speech perception. There are measurable effects of training and experience on speech processing in the auditory brainstem. This is consistent with an active model of speech perception in which attention and experience shape the earliest levels of sensory encoding of speech. Although current data do not exist to support online changes in this kind of processing, this is exactly the kind of prediction an active model of speech perception would make but is entirely unexpected from any of the current models of speech perception.

[This page is intentionally left Blank]

CHAPTER 6
SPOKEN WORD RECOGNITION

INTRODUCTION

Word is the minimal unit of the meaning. Understanding spoken language means we need to connect the sound that is issued with the meaning that is spoken. At the core of this process is the recognition of spoken words, it is the result of the representation of knowledge in the mental lexicon that serves as a true bridge or as a liaison between the sound issued with the meaning spoken, linking the phonological properties of specific word-forms to their syntactic and semantic attributes. The recognition of spoken words can be seen as a process of classifying the stimulus we hear as a form of words. In this paper, three fundamental functions of spoken word recognition will be discussed. Frauenfelder and Tyler (1987) classified the functions required of any theory of spoken word recognition into three stages. Initial contact is how input interfaces with and activates lexical representations. Selection describes how the set of activated lexical alternatives is evaluated with respect to the sensory input. Integration refers to how candidates are evaluated with respect to the linguistic and nonlinguistic context, in order to identify which is the likeliest candidate for recognition as well as to build larger linguistic structures. These three functional requirements have to be realized in some way in any model of spoken word recognition. This paper will also discuss various phenomena in spoken word recognition, as well as the process in spoken word recognition.

A. The Process of Spoken Word Recognition

1. Lexical Processing

In the following sections we will first outline what we consider to be the major phases involved in lexical processing and indicate how different theoretical positions have dealt with each of them. Then we will present a brief overview of the way in which context effects of different types have been assumed to intervene in these phases of lexical processing. Throughout this introduction we will raise some of the issues that continue to dominate research in lexical representation and process. One of our objectives is to confront the terminological confusion plaguing word recognition research. Indeed, it is clear, even from the contributions to this volume, that we are still far from a terminological consensus. Basic terms like "word recognition " and " lexical access" are often used to refer to very different processes (d . Tanenhaus & Lucas, 1987, this issue). We will attempt to identify in as theoretically neutral a fashion as possible the major aspects of lexical processing in an effort to develop a terminology which is consistent with various theoretical frameworks.

1) Initial lexical contact

The process of recognising a spoken word begins when the sensory input- or ,more precisely, some representation computed from this input- makes initial contact with the lexicon. In this initial contact phase, the listener takes the speech wave as input and generates the representation(s) which contact the internally stored form-based representations associated with each lexical entry. A major question concerns the nature of the representation which makes contact with the lexicon. This representation has important consequences not only for which lexical entries are initially contacted but also for when they are contacted.

When someone speaks, the linguistic content and speaker characteristics (e.g., physiology of the vocal tract, gender, regional origin, emotions, identity) simultaneously influence the acoustics of the resulting spoken output. Additional sources of variability include rate of elocution, prosodic prominence, and the phonetic context in which each word is pronounced. Nonetheless, listeners are able to recognize acoustically different stimuli as instances of the same word, thus extracting the similarity that exists between these different tokens, and perceiving them as members of the same category. How are words mentally represented to allow for this complex categorization? The traditional (and dominant) view assumes that people represent the form of words as categories that abstract away from variability. Drawing on linguistic theories, the mental representation of a word form is usually conceived as a sequence of phonemes (sometimes themselves decomposed into a bundle of contrastive features). Within this framework, the ease with which a given pronunciation is categorized as a taken of a given word is assumed to depend upon the degree to which its components have characteristics typically associated with the word’s phonemes. Speaker-specific information is often viewed as a source of noise which does not contribute to the process of identifying the linguistic units present in the signal.

- Contact representations

Many different contact representationsh ave been proposed to mediate this initial phase- ranging from temporally defined spectral templates (e.g., Klatt , 1980) to abstract linguistic units like phonemes (e.g., Pisoni & Luce, 1987, this issue) or syllables (e.g., Mehler, 1981). The properties of these representations have potential consequencesfo r the size of the initially contacted subset of the lexicon. The richer or more discriminative the information in the contact representation, the smaller the number of lexical entries initially contacted. To illustrate this point, we can contrast the effect of a phoneme-based representation with that of a robust feature representation (cf. Shipman & Zue, 1982) in which only six classes of phonemes are distinguished. In the former case, because the description of the input is much richer, it is more discriminative, and the size of the initially contacted set is smaller than in the latter case.

- When initial contact occurs

The amount of speech required to compute the contact representation determines the moment at which initial contact can occur. Clearly, the longer the stretch of speech signal that the system needs to accumulate to construct this representation, the more the initial contact is delayed. We can contrast models with potentially immediate contact such as the LAFS model (Klatt, 1980) in which the first 10 ms spectral template initiates a path to a lexical entry, with models in which there is a much longer "dead period" during which no contact is possible. Consistent with the latter type of proposal, it has been suggested that the first syllable of a word (Bradley & Forster, 1987, this issue) or the first 150 ms of a word (Marslen-Wilson, 1984; Salasoo & Pisoni, 1985; Tyler , 1984) needs to be analysed before contact can be made.

In some models the first contact with lexical entries is based upon some initial portion of a word (Cole & Jakimik, 1980, Marslen-Wilson & Welsh, 1978). In the "cohort model", for example, the "word-initial cohort" contains all of the words in a language matching some beginning portion of the input (Marslen-Wilson & Tyler , 1980; Marslen-Wilson & Welsh, 1978). This view, in which priority is given to temporally early information , can be contrasted with approaches in which information which is physically more salient- irrespective of its temporal location- is used to contact the lexicon. For instance, Grosjean and Gee (1987, this issue) claim that stretches of the signal that are particularly reliable (such as stressed syllables) establish the initially contacted subset of the lexicon. These approaches all share the assumption that there is a discrete stage of initial contact which delimits a subset of the lexicon.

- Advantagesa nd disadvantageso f discrete initial contact

The obvious advantage of discrete initial contact is that not all the entries in the lexicon need to be considered in subsequent phases of analysis. However, there are problems associated with the assumption that lexical entries are only ever considered if they are included in the initial subset of words matching the contact representation. For example, the intended word will never be located when the contact representation is misperceived. In order to reduce the risk of such unsuccesful initial contact, the contact representation has to be constrained. It has to be broad enough to ensure that the intended word is contacted, and yet specific enough so that only a minimal number of entries is contacted.

2. Activation

The lexical entries that match the contact representation to some criterial degree during the initial contact phase are assumed to change in state. In the absence of a theoretically neutral tend for this change, we will refer to it as “activation”. Theories differ in the claims they make concerning the factors that determine the relative status of activated words. For instance, the original version of the cohort theory proposed that all lexical entries matching the contact representation were equally activated and therefore had equal status. In the search model described by Bradley and Forster (1987, this issue), the relative status (the term "level of activation" is not appropriate for this model) of lexical entries at lexical contact depends upon properties of these entries themselves- in particular, upon their frequency of occurrence in the language. Lexical entries are ordered (within their respective subset or “bins") according to frequency. In other models, such as the current version of the cohort theory (Marslen-Wilson, 1987, this issue) and the Trace model, the degree of activation of a contacted lexical entry varies depending on both its goodness of fit with the contact representation(s) and its own internal specifications (e.g.,frequency of occurrence).

Theories differ on the patterns of activation that follow initial contact. More specifically, they differ in the theories of similarity they assume. The original (Marslen-Wilson & Welsh, 1978), revised (Marslen-Wilson, 1987, 1989) and distributed cohort models (Gaskell & Marslen-Wilson, 1997, 1999, 2002) place great emphasis on word onsets. The real-time constraints of the speech signal motivate an emphasis on optimal use of bottom-up information as it becomes available. Since a word's onset is heard first, it should determine which lexical items are first activated. Thus, in the original cohort model, the set of activated lexical alternatives was constrained to a *word-initial cohort* of items that matched perfectly the phonemic representation of the first approximately 150 ms of a word's onset. In light of evidence that a word might be recognized even when its first sounds are altered (for example, due to mispronunciation, cf. Cole, 1973), the revised and distributed cohort models abandon the strict, all-or-none match constraint. Instead, lexica representations are activated as a function of their similarity to a spoken word, with this similarity being continuously evaluated rather than limited to the initial portion of the spoken word. Nonetheless, the models' emphasis on real-time processing maintains a special status to the spoken word's initial sounds, as they contribute to the activation of some words, and thereby the interpretation of subsequent spoken material will be biased in favor of these words (see the discussion of *Selection* below for a full description of how these biases might be implemented). The neighborhood activation model (NAM; Luce, 1986; Luce, Pisoni, & Goldinger, 1990; Luce & Pisoni, 1998) differs from any instantiation of the cohort model by predicting activation of words that reflects their global similarity with the spoken word.

3. Selection

After initial contact and activation of a subset of the lexicon, accumulating sensory input continues to map onto this subset until the intended lexical entry is eventually selected. This selection phase has been described in various ways: As a process of differentiation (McClelland & Rumelhart, 1986), reduction (Marslen-Wilson, 1984) or search (Forster, 1976). In the Trace model, the differential activation of lexical entries provides the basis for selection. Through processes of activation and inhibition , one entry eventually emerges as the most activated relative to all other entries. In contrast to this approach, the original formulation of the cohort theory saw this as an all-or-none process. The internal specifications of lexical entries were assesseda gainst the sensory input and those which failed to match dropped out of the cohort. Thus, entries were either in or out of the cohort. A rather different approach is taken in the most recent version of the cohort theory (Marslen- Wilson, 1987, this issue) where lexical entries failing to match the input are not dropped from the cohort completely, but rather their level of activation starts to decay in the absence of further bottom-up support. In search models, the correct word is selected by a process which searches through the frequencyordered set of lexical entries (Bradley & Forster, 1987, this issue).

4. Word recognition

We will reserve the term word recognition for the end-point of the selection phase when a listener has determined which lexical entry was actually heard. An important objective in approaches which emphasise the temporal nature of the recognition process, has been to determine the word recognition point, that is, the precise moment in time at which a word is recognised. It is widely accepted that listeners generally recognize words, either in isolation or in context, before having heard them completely (Grosjean, 1980; Marslen-Wilson, 1984; Marslen-Wilson & Tyler , 1980). The exact recognition point of any given word depends upon a number of factors including its physical properties (e.g., length, stimulus quality), its intrinsic properties (frequency), the number and nature of other words in the lexicon that are physically similar to this word (i .e., its competitors or fellow cohort members) and the efficiency of the selection process. If the simplifying assumption is made that the acoustic signal is recognised sequentially, categorically and correctly as a sequence of discrete segments (e.g., phonemes or syllables) and that the selection process retains only those lexical entries matching this sequence, then it is possible to determine the recognition point for each word. In this case, a word's recognition point corresponds o its uniquenessp point that is, the point at which a word's initial sequence of segments is common to that word and no other. If, however, the analysis of the input proceeds in a probabilistic rather than categorical fashion, then a word is not necessarily recognised at the uniqueness point , but rather later at the moment the sensory input matches one single lexical candidate better than all others by some criterial amount (Marcus & Frauenfelder, 1985).

5. Lexical access

Lexical access is the process of entering the mental lexicon to retrieve information about words. The mental lexicon is the database containing all words in the mind of the language user. Lexical information can be, for instance, orthographic (spelling), phonological (sound), or semantic (meaning) in kind. Word recognition can then be defined as the process of retrieving these word characteristics on the basis of the input letter string. Although these different characteristics might become active under many circumstances (for instance, phonological codes may become available automatically), particular tasks may require specific kinds of lexical information to be performed. For instance, if one needs to decide whether a particular letter string is a word in the target language or a nonword (lexical decision), orthographic, phonological, and semantic information could in principle all be used. However, if one must name a presented word, the retrieval of its phonological information is indispensable to access the word's articulatory code. Finally, if asked to semantically categorize the object represented by the word (e.g., Is a hammer a tool?), the word's meaning information must be found before a response can be initiated.

The goal of lexical processing is to make available the stored knowledge associated with a word.(cf. Johnson-Laird , 1987, this issue) so that this can be used to develop a meaningful interpretation of an utterance. We use the term lexical access to refer to the point at which the various properties of stored lexical representations- phonological, syntactic, semantic, pragmatic- become available. One central question is when does this lexical information become available to the rest of the language processing system?

Most theories agree that some form-based information must be available in the initial contact phase of lexical processing- otherwise there would be no basis for a match with the sensory input. There is disagreement, however, on the point at which other types of stored lexical knowledge become available. The range of different views is exemplified by the contrast between the cohort and search models. In the cohort model, all stored information is activated simultaneously upon initial contact (Marslen- Wilson & Tyler, 1980). In the search model, although some form-based description must be made available early in the process (upon initial contact), stored syntactic and semantic information does not become available until a word is accessed and recognised (Forster, 1976, 1979). This is because such information is stored centrally. In a master file which is not entered until the word has been recognised (a process which takes place in the access "bin").

The assumed relationship between lexical access and word recognition variesdepending upon the theory. In models like that of Bradley and Forster (1987, this issue), lexical access and ~ord recognition, as defined here, are indistinguishable (although the authors themselves introduce another theoretical distinction between the two) since lexical information becomes available (lexical access) only when a single lexIcal entry has been found (word recognition). In models like the cohort model, there is a clear difference in that lexical access precedes word recognition. Up to now we have only discussed the phases involved in recognising words and accessing stored lexical information . What remains to be considered now is how higher-order context influences spoken word recognition.

B. Sentence and Context Effects

Swinney (1979) and Tanenhaus and colleagues (1979) reported classic results that motivated a modular view of form and meaning processing that dominated psycholinguistics for nearly 15 years. In a neutral context, ambiguous homophones prime all consistent meanings (e.g., BUG primes both *insect* and *spy*). A context biased towards one meaning eventually primes the context-appropriate meaning much more strongly than other meanings, but both Swinney and Tanenhaus and associates found that such effects appeared to take a few hundred milliseconds to emerge. Such results suggested two stages of processing: exhaustive bottom-up form access followed by semantic-based selection. Shillcock and Bard (1993) demonstrated that immediate effects *can* be observed with very strong expectations established by syntactic and semantic constraints. For example, given a context such as, “John said he didn’t want to do the job, but his brother would”, no cross-modal priming was found for timber (a semantic relative of wood) from would even when the probe was displayed *during* the critical word. Using the VWP, Dahan and Tanenhaus (2004) found that sentence contexts predicting a specific lexical item seemed to eliminate phonological competition from context-inappropriate competitors. Similarly, Magnuson, Tanenhaus, and Aslin (2008) found that syntactic and pragmatic constraints (whether an adjective or noun was expected next, given VWP display contents) immediately restricted competition; no phonological competition was observed from nouns when adjectives were expected and vice versa.

Chambers, Tanenhaus, and Magnuson (2004) found that object affordances and implications of a verb and/or instrument constrained competition. For example, given an instruction to “pour the egg . . .” when there were a liquid egg and an egg still in its shell, fixation proportions to the egg-in-shell were no greater than to unrelated objects. Given a hook to manipulate objects in a workspace and two whistles (only one of which was hookable, via a string attached to it), fi xations to the unhookable whistle did not differ from fi xation proportions to unrelated objects. Such results demonstrate *preemption*: the absence of competition expected from the bottom-up input, which is weak evidence for *anticipation* . Strong evidence for anticipation comes primarily from two sorts of studies: event-related potential experiments where large N400 responses indicate specifi c word expectations (e.g., at “an” in “The day was breezy, so the boy went outside to fl y *an* airplane”, given the very strong expectation for the fi nal noun phrase to be “a kite”; DeLong, Urbach, & Kutas, 2005) and VWP studies where expected items are fi xated before they are named. For example, given a display with a boy, a piece of cake, and some toys, fixations were equivalent for inanimate objects when subjects heard, “The boy will move the . . .”, but fi xations were directed toward the cake anticipatorily given, “The boy will eat the . . .” (Altmann & Kamide, 1999). Kamide, Altmann, and Haywood (2003) reported more complex interactions of scenes and word meaning. Given two possible riders (girl , man) and two rideable objects (carousel , motorbike), fi xation proportions favored expected relationships (greater fi xations to carousel than to motorbike given, “The girl will ride the . . .”, and vice versa for, “The man will ride the . . .”), although expectations appear to be probabilistic (e.g., although most fi xations were directed to carousel given, “The girl will ride the . . .”, fi xations to motorbike were greater than to non-ridable items). Ferretti, McRae, and Hatherell (2001) found additional support for strong impact of sentence context. After establishing that verbs prime typical agents, patients, instruments, and even specifi c features of patients (manipulating primes NAÏVE), though not typical locations, they presented auditory sentence fragments such as, “She arrested the . . .” (agent role fi lled, should only prime patient) or “She was arrested by the . . .” (patient role fi lled, should only prime agent) and then presented a visual target word appropriate for an agent or patient role (e.g., cop / crook), which participants had to name. Naming was facilitated only for expected roles. However, given that the VWP has proved more sensitive than a variety of cross-modal paradigms (e.g., Allopenna et al., 1998), Kukona, Fang, Aicher, Chen, and Magnuson (2011) explored similar sentence constraints using the VWP. In one experiment, every sentence was about something that “Toby” was doing (e.g., “Toby arrests the crook”). In critical displays, excellent agents and patients of the verb were displayed (e.g., crook and policemen).

Despite the fact that the agent role was *always* fi lled by Toby (always pictured in the center of the display), equivalent “anticipatory” fi xations were observed to both good patients and good agents (which would not be expected if participants make optimal use of context); fi xations reliably favored patients only after the onset of the word naming the patient. A second experiment demonstrated reliable anticipatory preference with additional syntactic cues and time for constraints to have impact; all sentences were about things that happened to Toby (e.g., “Toby was arrested by the . . .”), and, while initial fi xations to the good patient and agent were equivalent as the verb was heard, a reliable anticipatory preference to fi xate the agent emerged during the preposition. It seems that naming in the Ferretti and colleagues (2001) study measured the late dominance of the context-appropriate role and was not suffi ciently sensitive to pick up the weaker early coactivation of both roles.

Finally, meaning ascribed to objects in the world also includes something like discourse tags or situation models. Chambers and San Juan (2008) used the VWP and had participants follow a series of instructions with displayed items (e.g., “Move the chair to area 2; now move the chair to area 5; now return the chair to area 2”). An instruction beginning “now return” led to anticipatory eye movements to previously “visited” areas. Thus, recognition of spoken words entails accessing long-term knowledge of semantic features but also situation-specifi c mappings among words, the environment, and discourse history.

1. Word Frequency Effects

The word frequency effect is perhaps the most widely studied phenomenon in spoken word recognition. Although word counts have been almost universally obtained from written language (Kucera & Francis, 1 967 ; Thorndike & Lorge, 1944) , they are assumed to approximate the distribution of words in spoken language (see Carterette & Jones, 1 974). I High-frequency words are recognized faster and more accurately in almost all word recognition tasks . Savin (1963) provided one of the earliest demonstrations of the frequency effect in spoken word recognition. He presented subjects with high- and low-frequency words in white noise for perceptual identification and found that high-frequency words were recognized at lower signal-to-noise ratios than low-frequency words. Furthermore, when subjects made errors, they responded with words that were higher in frequency than the words that were actually presented . Using the gating task, Luce et al. (1984) reported that high-frequency words required less acoustic phonetic information for recognition than low-frequency words. P . A. Luce (1986) found that lexical decision times for spoken words were faster to high-frequency words than to low frequency words. Under certain experimental conditions, the phoneme monitoring task is also sensitive to word frequency; phonemes in high-frequency words are detected faster than phonemes in low-frequency words (Eimas et al . , 1 990; Lively & Pisoni, 1 990).

A number of mechanisms have been proposed to account for the word frequency effect. For the present discussion, however, two basic approaches have been proposed: One approach suggests that frequency effects are the result of a bias; the other approach suggests that frequency effects reflect the sensitivity of the word recognition system. In his review of word frequency effects, P. A. Luce (1986) cites a number of different models that incorporate frequency as biasing information. For example, Savin (1963) proposed that subjects engage in a sophisticated guessing strategy that favors high-frequency words when stimuli are presented in noise. Guessing among alternatives is mediated by the number and frequency of the items that must be discriminated (P. A. Luce, 1 986; Treisman, 1978a, 1978b) . Balota and Chumbley (1984) and Broadbent (1967) have also proposed other bias-related accounts of frequency effects . Each of these approaches involves manipulating the criteria used to select a response in order to favor high-frequency words over low-frequency words.

An alternative to the bias approach is the assumption that frequency effects arise from changes in the sensitivity of the word recognition system. According to this account, the thresholds or resting activation levels of lexical representations are set in a frequency-sensitive manner (McClelland & Rumelhart, 1981; Morton, 1969) . Thus , the selection criteria are the same for high- and lowfrequency words , but less stimulus information is required to recognize highfrequency words

2. Lexical Similarity Effects

While word frequency effects are assumed to reflect the statistical distribution of spoken words in the language , Lexical Similarity Effects reflect relationship among phonetically similar words . In this case, the focus is on how the recognition of a given word is affected by the presence of other phonetically similar words in the language . One measure of similarity that has been used in the spoken word recognition literature is an adaptation of Colt heart , Develaars , Jonasson, and Besner's (1976) N-metric . The N-metric was developed to assess similarity effects in visual word recognition. Two words are considered to be visual neighbors if they differ from each other by only one letter. The auditory analog of the N metric assumes that words are neighbors if they differ by only one phoneme (Greenberg & Jenkins, 1 964; Landauer & Streeter, 1 973). According to these definitions , sand and wand are visual neighbors, but not auditory neighbors ; vote and vogue are auditory neighbors , but not visual neighbors; bat and cat are auditory and visual neighbors . Words that are similar to many other words come from dense neighborhoods: words that are similar to few other words come from sparse neighborhoods. Although the N-metric is only a gross measure of perceptual similarity, its use has led to a number of important findings.

Several experimental paradigms have produced consistent findings based on manipulations of lexical similarity. For example, although Savin (1 963) did not explicitly manipulate perceptual similarity, he reported that misidentifications across subjects in perceptual identification were highly regular. The pattern of responses suggested that subjects were generating and selecting their responses from a set of phonetically similar alternatives . Luce found that iden tification accuracy was dependent on the number and frequency of neighbors of a stimulus word (P. A . Luce, 1 986; see also Pisoni, Nusbaum, Luce, & Slowiaczek, 1 985) . Moreover, Pisoni et al. (1985) found that lexical similarity plays an important role in determining identification accuracy when stimulus words were controlled for frequency. They reanalyzed data originally collected by Hood and Poole (1980) and found that words from dense neighborhoods were identified less accurately than words from sparse neighborhoods , although all words were of equal frequency.

Treisman (1978a, 1 978b) addressed the issue of lexical similarity and its effect on identification accuracy in his partial identification theory . He proposed that words can be represented as points in a multidimensional space and that identification accuracy is highly dependent on the discriminability of the items. As words become more discriminable, and thereby less similar to other words, identification accuracy increases. In the categorization literature , Krumhansl (1978) has also considered how identification of an object is affected by the presence of other similar objects. According to her distance-density model, density changes the discriminability of objects in the space; objects in dense regions require finer discriminations than equally similar objects in sparse regions.

The evidence presented so far suggests that similarity effects are strictly inhibitory. That is, words from dense neighborhoods tend to be identified less accurately than words from sparse neighborhoods . However, other findings indicate that the role of similarity may depend on the nature of the task and the modality of presentation. In tasks such as auditory perceptual identification and lexical decision, a unique item must be discriminated from its competitors before a response can be provided . When a large number of alternatives must be discriminated, responses tend to be slow and inaccurate. However, when only a gross discrimination is required, responses are faster to words from dense neighborhoods . For example, words from dense neighborhoods are responded to faster than words from sparse neighborhoods in phoneme categorization (Lively & Pisoni, 1 990). Similar advantages have been observed in the visual word recognition literature for words from dense neighborhoods . Andrews (1989, 1992) reported that naming and lexical decision responses to lowfrequency words from dense neighborhoods were faster than responses to lowfrequency words from sparse neighborhoods. An important debate has developed in the word recognition literature concerning lexical similarity effects . One position proposes that the number of words in a neighborhood has an impact on recognition. For example, Andrews (1989, 1 992) has shown advantages for words from dense neighborhoods in visual lexical decision and naming. A second position argues that similarity effects reflect the frequency of the items that must be discriminated. Grainger, O'Reagan, Jacobs , and Segui (1989; Grainger, 1 990; Grainger & Segui, 1

990) have shown that visual lexical decision and naming responses are slower when a stimulus word is similar to at least one word that is higher in frequency. A third position is that both the number of words that must be discriminated and their frequencies will have an impact on word recognition. P. A. Luce (1986) showed that both neighborhood size and neighborhood frequency accounted for significant and independent portions of the variance observed in auditory perceptual identification and lexical decision experiments.

C. Models of Spoken Word Recognition

Most accounts of spoken word recognition were originally derived from models of visual word recognition. The translation across modalities has not always demonstrated a concern for the basic problems of speech perception discussed earlier. Despite this oversight, three basic metaphors have been used to describe the processes involved in mapping an acoustic waveform onto a representation stored in long-term memory. One set of models assumes that words are retrieved from memory through a search process. In this case, the lexicon is endowed with special organizational properties that allow for the fast and efficient search through a large number of stored alternatives. The second class of models assumes that words are recognized through an activation process . In general, these models assume that lexical candidates are activated in proportion to their match to the incoming signal. The criteria for response selection vary from model to model, but generally items are selected on the basis of their relative activation levels. The third class of models combines assumptions from the first two types of models. These hybrid models assume that a number of candidates are initially activated and then a search process is engaged to find the proper entry . Examples of models from each class are considered here. We begin with the search model, followed by the pure activation models , and conclude with the hybrid models. Our goal is briefly to outline the design principles of each model and to point out their relative strengths and weaknesses.

1. Search Models

In its original form, Forster's autonomous search model (1976,1979) made strong claims about the structure of the linguistic processing system and the subsystem dedicated to visual word recognition. Although the model has been revised in recent years (Forster, 1 987, 1989), the original instantiation provides a pedagogically important example of an autonomous model of language processing. According to Forster, the word recognition system is divided into severalparts. One component, or peripheral access file , deals with information about how words are spelled; a second is devoted to acoustic phonetic input. At the earliest stage of processing, an unrecognized pattern is submitted to the proper peripheral access file for analysis . Words in each file are arranged in a number of separate bins. Items within each bin are assumed to be arranged in descending order of frequency. Thus, high-frequency words are searched before low-frequency words. When an entry in one of the peripheral bins or access files is matched to the input , a pointer to an entry in the master lexicon is retrieved. Once the pointer has been traced into the master lexicon, properties of the word , such as its syntactic function and semantic composition are made available to the General Problem Solver (GPS). Because Forster (1 976, 1 979) assumes that linguistic processes operate independently of each other, word recognition is not influenced by syntactic or semantic computations . Information about the plausibility of the incoming message is sent to the GPS from the cognitive system that is responsible for extralinguistic conceptual knowledge . The role ofthe GPS is to collect and integrate the output from each processor and to decide how to act on that information . It serves as a control mechanism or executive to oversee the operations of the entire processing system.

Over the years, the search model has undergone extensive revision (Forster, 1987, 1989). The revised version of the search model is now similar to activationbased models discussed below. In the original model, a single comparator matched the incoming signal to lexical representations in the peripheral access files . This created a problem in terms of the number of bins that needed to be searched, relative to the observed speed of word recognition (Feldman & Ballard, 1982; Forster, 1989). In the revised model, Forster proposed a separate comparator for each bin. The addition of multiple comparators solves the problem of determining which bin to search first. As Forster points out, a logical extension of the model is to dedicate a comparator for each lexical entry. This addition, however, would effectively transform the model from a search-based model to an activation-based model and would abandon the frequency-ordered search mechanism that made the original version so appealing. A second addition to the model has been to assume different levels of activity among lexical entries (Forster, 1 987) . This change was motivated by a number of form-based priming findings that showed facilitation due to orthographic similarities among primes and targets . Although Forster contends that his notion of activation is qualitative rather than quantitative, these assumptions make the new model very similar to activation models (see also Grainger & Segui, 1990).

2. Activation Models

1. Morton 's Logogen Model

Morton' s Logogen model (1 969, 1 970, 1 982) was one of the earliest examples of a direct access, activation-based model of word recognition. Each word in the lexicon is represented by an independent, passive detecting unit called a logogen that contains orthographic, phonological, syntactic, and semantic information . Logogens monitor the input signal (auditory or visual) for relevant information . As information is gathered, the activation levels of the logogens rise. When a logogen has gathered enough information to cross a recognition threshold, the information contained in the logogen becomes available to the cognitive system.

The logogen model has several attributes that allow it to account for phenomena in the word recognition literature. First, the response thresholds for individual logogens can be modified by word frequency: logogens corresponding to high-frequency words have lower thresholds . This mechanism allows the model to predict word frequency effects in a task such as auditory perceptual identification because less acoustic phonetic information is required to cross the thresholds of logogens corresponding to high-frequency words.

Second, the logogen model is interactive: expectations generated from the syntactic and semantic structure of the input affect logogens' activation levels. Thus, words that are syntactically consistent with the input require less acousticphonetic information for recognition than they would if they were presented in isolation. The interactive assumption also allows the model to account for semantic priming. Logogens that are semantically related to the prime are partially activated when the target is presented. Thus, when doctor is presented, the logogen corresponding to nurse also moves toward its threshold, given their semantic relation. This leads to the prediction that nurse should be recognized faster and/or more accurately than it would have been without the presentation of a prime.

Despite its simplicity, the logogen model has several problems. In its current instantiation , the model, like many others in the field of word recognition , is vague (Pisoni & Luce, 1 987) . The model is under specified in three ways : First, the model does not specify the perceptual units that are used to map acoustic phonetic input onto logogens in memory . Second, the model does not specify how different sources of linguistic information are integrated. Finally, the model cannot account for lexical similarity effects . Because logogens are activated independently, they are unable to facilitate or inhibit the activation levels of other phonetically similar logogens . Thus, effects of lexical density cannot be accommodated by the logogen model in its present form.

2. Marslen- Wilson 's Cohort Theory

Cohort Theory (Marslen-Wilson, 1 987, 1 990; Marslen-Wilson & Welsh, 1978 ; Marslen-Wilson & Tyler, 1 980) is example of an activation-based model that was directly motivated by several observations about spoken word recognition. First, spoken word recognition may be characterized by early selection of several hypothesized lexical candidates . Marslen-Wilson (1 987) defined early selection as "the reliable identification of spoken words, in utterance contexts, before sufficient acoustic-phonetic information has become available to allow correct recognition on that basis alone" . Second, listeners recognize words very quickly, in close to real time. Marslen-Wilson (1985, 1987) reported that word recognition occurs within 200-250 ms of the beginning of the word . Finally, spoken word recognition is a highly efficient process . Marslen-Wilson (1984 , 1987, 1990) argues that listeners are very sensitive to the recognition point of a word-that is, the point at which the word diverges from all other possible candidates. Thus, only a minimal amount of acoustic-phonetic information is needed for accurate identification. As a consequence of these observations, Marslen-Wilson and his colleagues have proposed and developed an influential model that operates via a contingent process. According to cohort theory , spoken word recognition involves deciding what was presented and what was not. Thus, the recognition process involves both identification and discrimination (see P. A. Luce, 1 986).

Marslen-Wilson (1 987) divided spoken word recognition into three subprocesses. At the first level, the word recognition system makes contact with a low-level, acoustic-phonetic representation of the input signal. A set of lexical candidates is activated during this stage. Next, a selection process is used to choose a single item from the word-initial cohort. Finally, the highest level subprocessor takes the selected lexical item and integrates it into the available syntactic/semantic discourse. For the purposes of this chapter, the access and selection mechanisms are of the most interest. Cohort theory assumes that the lexicon is initially contacted in a strictly bottom-up, data-driven manner. Abstract lexical representations stored in memory are accessed via acoustic-phonetic representations generated from the input signal.

The lexical candidates generated by the access mechanism form the word-initial cohort . In the original version of the model , MarslenWilson and Welsh (1978) argued that each member of the word-initial cohort was assumed to have the same initial phoneme . In more recent instantiations of the model, however, the word-initial cohort for a word is assumed to contain words that have phonetically similar initial phonemes , thereby loosening the constraints on cohort membership (Marslen-Wilson, 1 987) . After access, the word-initial cohort is then submitted to the selection mechanism. Unlike the access mechanism, the selection mechanism is sensitive to multiple sources of information, such as acoustic phonetic input, word frequency, and syntactic/ semantic context. Marslen-Wilson (1 990) suggested that the activation levels of items in the word-initial cohort change according to their overall consistency with the incoming signal . Items that are consistent with the input remain strongly activated, while the activation levels of inconsistent items drop off.

An important change has been made in the selection phase of the most recent version of cohort theory (Marslen-Wilson, 1 990). Continuous activation functions have been added to the model. In early instantiations of the model, candidates were assumed to be either in or out of the cohort in a binary manner (Marslen-Wilson & Tyler, 1 980; Marslen-Wilson & Welsh, 1 978) . With continuous activation functions, candidates can be judged relative to competitors . The change from binary to continuous activation functions now allows the model to account for word frequency effects. Marslen-Wilson (1 987, 1 990) assumes that high-frequency words receive more activation per unit of information than low-frequency words. This assumption predicts that low-frequency words will be harder to recognize in the context of high-frequency words because the activation levels of high-frequency words dominate , even after they are no longer consistent with the acoustic-phonetic input or the syntactic/ semantic context. Thus, low-frequency words with high-frequency competitors may be recognized more slowly or less accurately due to the high activation levels of similar high-frequency words. P. A. Luce (1 986) has made the same point concerning the importance of the relative frequencies of competitors in the recognition process (see also Grainger, 1 990) . In evaluating cohort theory as a model of spoken word recognition, two desirable properties are apparent. First, the model is explicit concerning the time course of spoken word recognition. Unlike models derived from visual word recognition, cohort theory acknowledges the temporal nature of speech. The concept of the recognition point of a word gives the model considerable power in determining when a word can be discriminated from other lexical candidates and helps to account for the real-time nature of spoken word recognition.

Second, the model acknowledges the importance of the beginnings of words and the left-to-right nature of spoken word recognition (Cole & Jakimik, 1980 ; Pisoni & Luce, 1 987) . For example, Marslen-Wilson and Welsh (1 978) found that phonemes near the ends of words were restored more often during shadowing than phonemes at the beginnings of words. Similarly , Cole, Jakimik, and Cooper (1 978) found that mispronunciations in word beginnings were detected more accurately than in word endings. The importance attached to word-initial information in cohort theory is also a problem with the model. It is difficult to understand how listeners recover from errors caused by activating the wrong word-initial cohort. Marslen-Wilson (1987) has attempted to make the model more flexible by assuming that items are represented featurally, rather than phonemically. In this case, all items in the new word-initial cohort share similar initial phonemes , but all the initial phonemes need not be the same. With this softening of assumptions, however, some mechanism needs to be provided to explain how the acoustic signal is transformed into phonetic features. And, it should be pointed out that recognition points are defined by left-to-right segmental structure, not a featural similarity metric. Although a bottom-up access phase is still plausible, the access mechanism needs to explain how the acoustic signal is mapped onto what is essentially an idealized symbolic , featural representation of speech.

3. McClelland and Elman's Trace Model

The trace model of speech perception and spoken word recognition (Elman & McClelland, 1 986; McClelland, 1 99 1 ; McClelland & Elman, 1 986) was designed to deal explicitly with the problems of variability, linearity, and segmentation. It is a connectionist model, based on McClelland and Rumelhart' s interactive activation model of visual word recognition (1 98 1 ; Rumelhart & McClelland, 1 982) . Individual processing units, or nodes, are dedicated to features, phonemes, and words . Nodes at each level of representation are highly interconnected. Feature nodes are connected to phoneme nodes, and phoneme nodes are connected to word nodes . In addition to between-level connections, units at the same level are also interconnected . Connections between levels are strictly facilitatory , symmetric, and bidirectional; connections within levels are inhibitory . This assumption lends interactivity to the model because higher level lexical information can influence the activation of phonemes at lower levels . Nodes influence each other in proportion to their activation levels and the strengths of their interconnections.

Unlike the logogen model, trace provides an explicit description of the time course of speech perception and spoken word recognition. The concept of a trace, which represents the working memory of the model , comes from assumptions about excitatory and inhibitory connections among processing units and temporal distribution of inputs . Activation levels of consistent units are increased as inputs are presented to the model, due to the excitatory connections between layers of nodes : activation levels among competing nodes are inhibited in proportion to their degree of overlap. As excitation and inhibition are passed among the network' s nodes, a pattern of activation, or trace, is developed to represent the processing history of the input . The selection of a unique item for recognition from the activated candidates is governed by R. D. Luce' s (1959) choice rule.

Trace has several appealing properties as a model of speech perception and spoken word recognition. First, it is a dynamic model because the recognition process is viewed as a temporally extended event. Unlike logogen, trace gives an explicit account of the unfolding of spoken word recognition. Second, Trace ' s use of bidirectional connections between levels of representation gives an explicit account of how data-driven and conceptually driven processes interact. Third, because of its architecture, trace deals with problems unique to spoken word recognition, such as linearity and segmentation . Fourth, trace can account for a number of phenomena in the speech perception and spoken word recognition literature without making ad hoc assumptions . For example, trace shows boundary shifts in categorical perception by using high-level lexical information to resolve ambiguous phoneme segments in favor of those that create words (Ganong, 1 980). This falls naturally out of the interactive activation architecture, with no additional assumptions required.

In spite of the model' s strong points, two serious weaknesses of trace need to be considered . First, the model ' s treatment of time is unrealistic because each phoneme is assumed to have the same duration . Klatt (1 989) has pointed out that this assumption ignores the inherent temporal variability of speech. Second, trace has only a very small lexicon of monosyllabic words. It is unclear how the model will perform if it is scaled up to a more realistic estimate of the size of the adult lexicon.

Trace makes an intermediate prediction: It activates both onset- and rhyme-overlapping words, because, as in the neighborhood model, words can be activated even if they mismatch at onset. However, unlike the neighborhood model, trace represents time: Words that become activated early in the spoken input have an advantage over words that become activated later, because more of the spoken word has been heard and selection mechanisms are then more effective at favoring the best matching candidate. Thus, trace predicts activation of both onset- and rhyme-overlapping candidates, although at different times and of different amplitude. Allopenna, Magnuson, and Tanenhaus (1998) provided behavioral data supporting this prediction. They estimated lexical activation to word candidates by monitoring eye movements to pictures as participants followed verbal instructions to move an item on a computer screen. Fixations were closely time-locked to the speech (with a lag only slightly larger than that required to plan and launch an eye movement), and mapped closely onto phonetic similarity over time (with higher and earlier fixation proportions to onset overlapping competitor than rhyme-overlapping competitor) as well as response probabilities generated by trace. This study highlights the importance of a measure of lexical activation over time, given the rapid evolution of lexical activation as the spoken input is heard. The Allopenna et al. (1998) study highlights one shortcoming of the similarity model embodied in NAM to the study of spoken word recognition.

The temporal distribution of similarity is not considered; *dab* and *bad* are assumed to be equally active upon hearing *dad* (ignoring frequency for the sake of the example). NAM fails to capture the temporal dimension of speech and the special status that the initial sounds have due to their temporal precedence (Marslen-Wilson & Zwitserlood, 1989). It also gives too much weight to the match in the number of segments or syllabic structure by entirely excluding the contribution of words that are more than one phoneme longer than the word to be recognized, despite evidence suggesting that words of different lengths affect the processing of a given word (Marslen-Wilson, 1984, 1987). The algorithm cannot be easily extended to the computation of competition environment for polysyllabic words, as most of these words have very few, if any, competitors under the one-phoneme difference definition. Finally, the one-phoneme shortcut metric, which has been most widely used by researchers and has proven useful in stimulus selection and experimental control, treats any phoneme deviation equally, regardless of its phonetic nature. Confusion between two words differing by one-phoneme addition or substitution, or confusion between two words differing by a vowel or a consonant, are all assumed to be equivalent, despite empirical evidence that the nature of the phonetic feature(s) that differ between two words is an important factor in accounting for word confusions (e.g., Bailey & Hahn, 2005; Hahn & Bailey, 2005; see also van Ooijen, 1996).

4. Luce's Neighborhood Activation Model

P. A . Luce ' s neighborhood activation model (NAM; Goldinger et aI., 989; P. A. Luce, 1986; Luce, Pisoni, & Goldinger, 1990) is an activationbased processing instantiation of R. D. Luce' s (1 959) biased choice rule . The fundamental principle underlying the model is that both frequency and perceptual similarity affect recognition . According to NAM, the frequencies of items in a word's similarity neighborhood have an important impact on recognition. The importance of neighborhood size and structure has been formalized using R. D. Luce's (1 959) biased choice rule [Equation (1)] . As applied in NAM, the rule predicts that the probability of correctly identifying a stimulus word Si is a frequency-weighted function of the probability that the stimulus word was presented versus the frequency-weighted probabilities that other similar words N ij were presented. Similarity is estimated as a measure of confusability among consonant-vowel and vowel-consonant sequences that are independent of any word or its neighbors. It is quantified in Eq. (1) by the conditional probability terms, which represent the conditional probabilities of identifying segments in the stimulus word . Viewed this way, the similarity between two words is a conditional probability of confusing one sequence of consonant- vowel and vowel-consonant combinations with another.

D. Functional Parallelism in Spoken Word Recognition

The overall process of spoken word-recognition breaks down into three fundamental functions. These I will refer to as the access, the selection, and the integration functions. The first of these, the access function, concerns the relationship of the recognition process to the sensory input. The system must provide the basis for a mapping of the speech signal onto the representations of word-forms in the mental lexicon. Assuming some sort of acoustic-phonetic analysis of the speech input, it is a representation of the input in these terms that is projected onto the mental lexicon. The integration function, conversely, concerns the relationship of the recognition process to the higher-level representation of the utterance. In order to complete the recognition process, the system must provide the basis for the integration, into this higher level of representation, of the syntactic and semantic information associated with the word that is being recognised. Finally, and mediating between access and integration, there is the selection function. In addition to accessing word-forms from the sensory input, the system must also discriminate between them, selecting the word-form that best matches the available input. These three functional requirements have to be realised in some way in any model of spoken word-recognition. They need to be translated into claims about the kinds of processes that subserve these functions, and about the processing relations between them during the recognition of a word.

CONCLUSION

While early traces of connections among word forms and meanings are evident in the first year of life, learning a word is a gradual process. Much of this learning process occurs underground: Representations of the forms and meanings of words are built up gradually before they begin to surface in children’s productive vocabularies. Perhaps the best evidence for this graded, online view of early language understanding comes from measures like eyetracking and event-related potentials. Unlike traditional looking-time or offline measures, these online methods allow for the detailed characterization of individual infants, allowing researchers to track the development of "partial knowledge." In addition, going forward, these methods will provide the best chance for researchers to test quantitative as well as qualitative models of development. This gradualist view contrasts dramatically with the binary construal of word learning as a process of "fast mapping" between word forms and their meanings, a metaphor that is often used along with statistics about the exponential growth of vocabulary. As demonstrated by the original work on "fast mapping," only a few exposures may be necessary for a veteran word learner to form a partial representation of a word (Carey and Bartlett, 1978). However, much more practice is necessary before the same learner can successfully interpret and produce that word appropriately across a range of contexts. In this review, we have attempted to give an overview both of the developmental progression in skills involved in word recognition by the young language learner, and in the historical progression of research on early spoken word recognition. In many ways, these two trajectories mirror one another, proceeding from the early groundwork laid by studies of auditory language processing and speech perception to a more complete understanding of the complexities involved in learning to communicate using words.

In recent reviews of the word recognition literature, Marslen-Wilson and Forster have commented that current models of spoken word recognition are becoming increasingly similar to each other. In their original forms, many of the models considered in earlier sections made strong predictions about the effects of different lexical variables on speed and accuracy . As each of these proposals failed various empirical tests , the models were revised and tested again. Many of the core assumptions that differentiated the various models have been weakened or lost, and this revision process has led to a convergence of the models on a few basic principles. Below, we consider several general themes that are common among contemporary models of spoken word recognition.

In summary, future progress in spoken word recognition will require taking a somewhat broader perspective than previous theoretical accounts of speech perception and spoken word recognition. In its current state, the spoken word recognition literature is filled with a number of very similar models, all containing slight variants of the same sets of ideas and principles. It is appropriate to reiterate Forster's concern that , in the future, researchers working on spoken word recognition will need to focus on developing a set of design principles for the next generation of models , rather than trying to decide whose interpretation of a particular model is the best. We believe that new models of spoken word recognition will be strengthened substantially by relating them directly to many of the current issues and theoretical efforts in learning, memory, and categorization.

[This page is intentionally left Blank]

CHAPTER 7
VISUAL WORD RECOGNITION: THE JOURNEY FROM FEATURES TO MEANING

INTRODUCTION

It has been more than a decade since the literature on visual word recognition has been reviewed for the Gernsbacher (1994) Handbook of Psycholinguistics, and there continues to be considerable interest in understanding the processes tied to the “word.” Understanding the journey from features to meaning has clearly made progress, but there is still some distance to go. In the present chapter, we will provide an update on the major issues that were covered in the 1994 chapter, and introduce many new issues that have arisen during the interim. This chapter will focus on isolated visual word recognition research; other chapters in this volume are devoted to auditory word recognition, and recognizing words in sentential context. The goal of the present review is not to provide in-depth reviews of every area addressed by word recognition researchers. This would far exceed space limitations. Rather, we will attempt to acquaint the reader with the richness and diversity of the empirical and theoretical issues that have been uncovered in this literature. The organization of the chapter is as follows: First, we will briefly outline why word recognition research has been central to a number of quite distinct developments in cognitive psychology, psycholinguistics, and cognitive neuroscience. Second, we will review the evidence regarding letter recognition, sub lexical organization, and lexical-level influences on word recognition. Interspersed within each of these sections is a discussion of some of the current theoretical developments and controversies. Third, we will review the literature on context and priming effects in word recognition; again, highlighting major theoretical developments and controversies. Fourth, we will discuss some limitations regarding inferences that are possible based on the available data, and highlight some recent developments that have provided additional leverage on such issues. The lexical unit is ideally suited for such work because words can be analyzed at many different levels, e.g., features, letters, graphemes, phonemes, morphemes, semantics, among others. As we shall see below, much of the work in visual word recognition has been devoted to identifying the functional roles of these different levels

In order to provide a framework for understanding the breadth of word recognition research, it is useful to list a few of the basic research issues that the word recognition literature has touched upon. Second, word recognition research has been central in the development of theories of automatic and attentional processes (e.g., Healy & Drewnowski, 1983; LaBerge & Samuels, 1974; Neely, 1977; Posner & Snyder, 1975). Third, word recognition research has also been central to developments regarding basic pattern recognition processes. One of the most difficult problems in pattern recognition research has been identifying the underlying subordinate critical features of a given pattern (e.g., Neisser, 1967). Finally, because words are relatively well-characterized patterns, they have been the focus of development of formal mathematical models of pattern recognition. For example, one of the first formal models in cognitive psychology was the Selfridge and Neisser (1960) Pandemonium model of letter recognition.

A. Features, Letters, and Modeling Constraints

We shall now review some of the variables that have been pursued in word recognition research. First, we shall attempt to break the word down into smaller, more tractable bits. Second, we will discuss work that addresses how orthography maps onto phonology in English. Third, we will discuss the influence of variables that can be quantified at the whole word level, e.g., frequency, familiarity, age of acquisition, orthographic neighborhood size, along with a set of additional semantic variables. Fourth, we will provide an overview of the influence of single word context on isolated word recognition, via a review of the priming literature. Sprinkled within each of these sections will be discussion of the major theoretical models and issues.

1. Features

A common approach to understanding pattern recognition is that a given pattern must first be broken down into features that are common to the set of patterns that one is interested in modeling. Some of the initial work in this area was developed by Gibson and Gibson (1955), who forcefully argued that feature-level analyses were an essential aspect of pattern recognition and, more generally, perceptual learning. These primitive features were the building blocks for pattern recognition. This provided researchers with a well-specified problem: what are the primitive features used in letter recognition? The hunt was on!

Fortunately, the feature analytic approach is ideally suited for letter recognition. Although there are differences across fonts, English orthography can be relatively well described by a limited set of features, such as horizontal lines, vertical lines, closed curves, open curves, intersections, cyclic redundancy, and others (see, for example, Gibson, Osser, Schiff, & Smith, 1963). Once researchers proposed such primitive features, both behavioral and neurological evidence began to accumulate that documented the role of such features in visual perception. On the behavioral side, there were studies of confusion matrices indicating that letters that shared features were more likely to be confused in degraded perceptual conditions, compared to letters that did not share many features (e.g., Kinney, Marsetta, & Showman, 1966). In addition, visual search studies by Neisser (1967), among others, indicated that subjects were relatively faster to find a given target letter (e.g., Z) when it was embedded in a set of letters that did not share many features with the target (e.g., O, J, U, D), compared to a set of letters that did share many features with the target (e.g., F, N, K, X). There was also exciting evidence accumulating during the same period that appeared to identify neural substrates that might sub serve feature-like detection processes. Consider, for example, the pioneering (and Nobel Prize winning) work by Hubel and Wiesel (1962, 1968). These researchers used single cell recording techniques to investigate neural activity in areas of the striate cortex in alert cats. When different stimuli were presented to the retina of the cat, there were increases in neural activity in specific cortical areas. Hubel and Wiesel found evidence that there were cells that appeared to be especially sensitive to visual stimuli that mapped onto such things as vertical lines, horizontal lines, angles, and even motion.

The importance of this work is very simple: it provided the neurological evidence that converged with the notion that pattern recognition ultimately depends upon primitive feature analytic processes. More recent work by Petersen, Fox, Snyder, and Raichle (1990) using positron emission tomography has extended this work to humans in demonstrating significant blood flow changes in specific areas of the striate cortex corresponding to feature-like detection systems for letter fonts in humans. At the same time behavioral and neural evidence was accumulating in support of features being used in pattern recognition, one of the first computational models of pattern recognition was developed. This was a model of letter recognition developed by Selfridge (1959; Selfridge and Neisser, 1960). The model initially coded the stimulus into a set of 28 visual features that provided support for the letters that were most consistent with those features. The Pandemonium model had the capacity to learn which features were especially discriminating among letters, and adjusted the weights for these features accordingly. As we shall see, the Pandemonium model predates by some 20 years important developments in letter and word recognition models.

It is quite amazing that the Pandemonium model worked so well given the computational hardware limitations in the late 1950s and early 1960s. Although most models of word recognition assume a first step of primitive feature identification, there are still many unanswered questions in this initial stage of processing: First, what is the glue that puts the features together? Specifically, once vertical lines, horizontal lines, and intersections have been detected, how does one put the features together to identify the letter T? We typically do not perceive free-floating features (for a review of the binding issue, see Treisman, 1996, and 1999). Second, what happens in the feature analytic models when distortions occur that modify the feature, e.g., does a 15° rotated vertical line still activate the vertical line detector? Third, and along the same lines, what are the critical features when the letters are distorted via different fonts or a novel style of handwriting?

Reading still proceeds in an acceptable fashion even though there are considerable changes in the critical set of features (see Manso de Zuniga, Humphreys, & Evett, 1991). Interestingly, there is some evidence that there may be differences in the way people process printed words and cursive handwriting. For example, case mixing disrupts reading performance with printed words (Mayall, Humphreys, & Olson, 1997) but can actually facilitate performance with handwriting (Schomaker & Segers, 1999). This suggests that distinctive word contours are more critical in handwriting recognition (for a description of a computational model of handwriting, see Schomaker & Van Galen, 1996). Fourth, are features across letters coded serially in reading, e.g., from left to right in English orthography, or is there a parallel coding of features? Based on the work by Treisman (1986), one might expect that there is an early parallel coding of features that is followed by a more capacity demanding binding process (see, however, Shulman, 1990). As we will see, the distinction between parallel and serial processing in word recognition has been a central area of debate in the literature (for a recent discussion, see Rastle & Coltheart, 2006). Finally, are features within letters the critical level of analysis in word recognition or are there supraletter and/or even word-level features (e.g., Purcell, Stanovich, & Spector, 1978) that are more important? Although there has been considerable progress in understanding how features contribute to pattern recognition (for a review, see Quinlan, 2003), there are still many questions that need to be resolved in mapping features onto letters. In lieu of getting bogged down in some of the more important fundamental aspects of visual perception, let us take the leap of faith and assume we have made it to the letter. Surely, things must get a bit more tractable there.

2. Letters

Assuming that features play a role in letter recognition, and letters are crucial in word recognition, one might ask what variables are important in letter recognition. For example, does the frequency of a given letter in print influence its perceptibility? Fortunately, there seems to be a relatively straightforward answer to this question. Appelman and Mayzner (1981) reviewed 58 studies that entailed 800,000 observations from a variety of paradigms that spanned 100 years of research. The conclusion from their review is very straightforward: letter frequency does appear to influence speeded tasks such as letter matching, naming, and classification tasks (e.g., is the letter a vowel or a consonant?). However, letter frequency does not appear to influence accuracy in perceptual identification tasks. The results from the Appelman and Mayzner study are intriguing for three reasons: First, a priori, one would clearly expect that frequency of any operation (perceptual, cognitive, or motoric) should influence performance, and hence, it is unclear why there is not a letter frequency effect in identification tasks. Second, as we shall see below, there is a consistent word level frequency effect in both response latency tasks and perceptual identification tasks, and hence, there at least appears to be a difference between frequency effects at different levels within the processing system, i.e., letters vs. words. Third, this is our first exposure of a general theme that runs across the word recognition literature, i.e., different tasks or analyses yield different patterns of data, and so it is incumbent upon the researcher to build a task of not only the targeted dimensions in word processing, but also the tasks that are used to tap these dimensions.

3. Features, Letters, and Word Interactions: Some Initial Models

Important theoretical issues regarding letter recognition date back to questions that were originally posed by Cattell (1885). The interest here is to define the perceptual unit in word recognition. A priori, it would seem obvious that the letter should be the primary unit of analysis in visual word recognition, i.e., words are made up of letters. However, Cattell (1885, 1886) reported evidence that was initially viewed as inconsistent with this notion. Cattell found that some words can be named more quickly than single letters. The problem this finding posed was very simple: how could the letter be the critical unit of analysis in word recognition, if words could be named more quickly than the letters that presumably make up the words? Along with the Cattell results, it was also reported that the exposure duration necessary to identify a word was in some cases less than the exposure duration necessary to identify a single letter. In fact, Erdmann and Dodge (1898) reported that the exposure duration necessary to identify four to five letters in a display was sufficient to read single words that could contain as many as 22 letters. Again, the conundrum is that if words can be better perceived than letters then how can letters be the basic unit of perception, since words are made up of letters?

Of course, an alternative account of this pattern of data is simply that subjects can use any available information regarding orthographic redundancy and lexical-level information to facilitate word processing, and such information is unavailable when single letters are presented. For example, if you thought you saw the letters T and H at the beginning of a short briefly presented word and the letter T at the end then you are likely to guess that there was the letter A between the TH and T, producing the word THAT. This was labeled the sophisticated guessing account of some of the initial findings. However, because of a seminal study by Reicher (1969), it appeared that there was more to this phenomena than simply sophisticated guessing. In Reicher's study, on each trial, one of three stimuli was briefly flashed (e.g., a single letter, K, a word, WORK, or a nonword, OWRK), after which a patterned mask was presented. After the mask was presented, subjects were presented with two letters (e.g., D and K) adjacent to the position of the previous target letter for a forced-choice decision. The remarkable finding here is that subjects produced reliably higher accuracy when the first stimulus was a word than when it was a single letter or a nonword. Because both the letters D and K produce acceptable words within the WOR context, subjects could not rely on pre-existing lexical knowledge to bias their response one way or the other (for an alternative view, see Krueger & Shapiro, 1979; Massaro, 1979). Hence, it appeared that subjects actually see letters better when embedded in words than when embedded in nonwords. This finding was termed the word-superiority effect and was also reported in a study by Wheeler (1970), so it sometimes also is referred to as the Reicher-Wheeler effect.

There were two important subsequent findings that constrained the interpretation of the word superiority effect. First, the effect primarily appears under conditions of patterned masking (masks that involve letter-like features) and does not occur under energy masking (masks that involve high-luminance contrasts, e.g., Johnston & McClelland, 1973; Juola, Leavitt, & Choe, 1974). In fact, it appears that the interfering effect of the mask is primarily on performance in the letter alone condition and does not produce much of a breakdown in the word condition (Bjork & Estes, 1973). Second, letters are also better recognized when presented in pronounceable nonwords (e.g., MAVE), compared to unpronounceable nonwords or alone (e.g., Carr, Davidson, & Hawkins, 1978; McClelland & Johnston, 1977). Thus, the word-superiority effect does not simply reflect a word-level effect.

The importance of the word-superiority effect derives not only from the information that it provides about letter and word recognition, but also from its historical impact on the level of modeling that researchers began to use to influence their theory development. Specifically, this effect led to the development of a quantitative model of word and letter recognition developed by McClelland and Rumelhart (1981; Rumelhart & McClelland, 1982; also see Paap, Newsome, McDonald, & Schvaneveldt, 1982). As noted earlier, this type of modeling endeavor set the stage for the explosion of interest in connectionist models of cognitive processes (e.g., McClelland & Rumelhart, 1986; Rumelhart & McClelland, 1986; Seidenberg & McClelland, 1989).

Figure 1 provides an overview of the architecture of the McClelland and Rumelhart (1981) model. Here, one can see the three basic processing levels; feature detectors, letter detectors, and word detectors. These levels are attached by facilitatory (arrowed lines) and/or inhibitory (knobbed lines) pathways. As shown in Figure 1, there are inhibitory connections within the word level and within the letter level. Very simply, when a stimulus is presented, the flow of activation is from the feature level to the letter level and eventually onto the word level. As time passes, the letter-level representations can be reinforced, via the facilitatory pathways, by the word-level representations and vice versa. Also, as time passes, within both the letter and word level representations, inhibition from highly activated representations will decrease the activation at less activated representations, via the within-level inhibitory pathways.

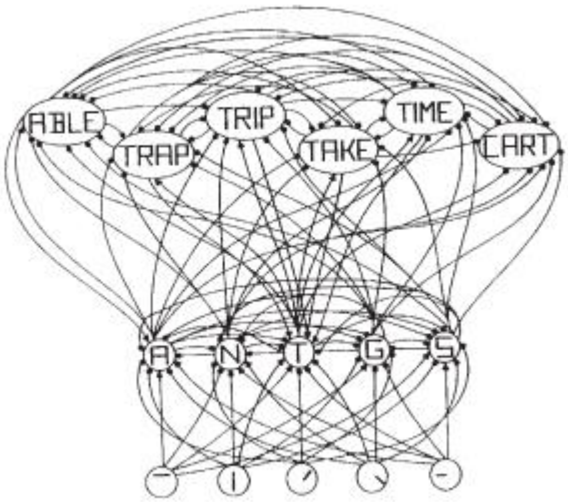


Figure 1. McClelland and Rumelhart's (1981) interactive activation model of letter recognition.

How does the model account for the word-superiority effect? The account rests heavily on the notion of cascade processes in the information processing system (see Abrams & Balota, 1991; Ashby, 1982; McClelland, 1979). Specifically, a given representation does not necessarily need to reach some response threshold before activation patterns can influence other representations, but rather, there is a relatively continuous transferal of activation and inhibition across and within levels as the stimulus is processed. Consider the letter alone condition in the Reicher paradigm, described earlier. When a letter is presented, it activates the set of features that are consistent with that letter. These feature detectors produce activation for the letter detectors that are consistent with those features, and inhibition for the letter detectors that are inconsistent with those features. Although there is some activation for words that are consistent with the letter and some inhibition for words that are inconsistent with the letter, this effect is relatively small because there is little influence of a single letter producing activation at the word level. Now, consider the condition wherein the letter is embedded in a word context. In a word context, there is now sufficient partial information from a set of letters to influence word-level activation patterns and this will produce a significant top-down influence onto letter-level representations, i.e., increase activation for consistent letters and decrease activation for the inconsistent letters. It is this higher-level activation and inhibition that overrides the deleterious influence of the patterned mask.

In passing, it is worth noting here that there is also evidence by Schendel and Shaw (1976) that suggests that features (e.g., lines) are better detected when the features are part of a letter than when presented alone. Hence, it is possible that there is also a letter superiority effect. Such an effect would appear to be easily accommodated within the McClelland and Rumelhart-type architecture by assuming that there are also top-down influences from the letter level to the feature level.

Interestingly, there is another phenomenon called the pseudo word superiority effect that would at first glance appear to be problematic for the McClelland and Rumelhart model. Specifically, letters are also better detected when embedded in pronounceable nonwords than when embedded in unpronounceable nonwords (Baron & Thurston, 1973; Carr et al., 1978), or presented in isolation (e.g., Carr et al., 1978; McClelland & Johnston, 1977). However, the interactive activation model can also accommodate this effect. Specifically, when letters are embedded in pronounceable nonwords, it is likely that there will be some overlap of spelling patterns between the pseudoword and acceptable lexical entries. For example, the pronounceable nonword MAVE activates 16 different four-letter words that share at least two letters within the McClelland and Rumelhart network. Thus, the influence of orthographic regularity appears to naturally fall out of the interaction across multiple lexical entries that share similar spelling patterns within the language. As we shall see below, the influence of orthographic regularity on word recognition performance has been central to many of the recent developments in word recognition research.

Although some orthographic regularity effects appear to naturally fall from this model, there are some additional intriguing insights from the model regarding other orthographic regularity effects. Consider, for example, the impact of bigram frequency. For example, the vowel pair EE occurs in many more words than the cluster OE. The available evidence indicates that there is relatively little impact of bigram frequency on letter recognition within a Reicher-type paradigm (Manelis, 1974; McClelland & Johnston, 1977; Spoehr & Smith, 1975). McClelland and Rumelhart have successfully simulated this finding within their interactive activation framework. Although high-frequency letter clusters are more likely than low-frequency letter clusters to activate many word-level representations, this activation will be compensated by the fact that there will also be more word-level inhibition across those activated representations. Because, as noted above, there are influences of the number of lexical representations that share more than two letters, the lack of an influence of bigram frequency would appear to indicate that there may be a critical limit in the amount of overlap across lexical representations that is necessary to overcome the deleterious effects of within-level inhibition. (Bigram frequency also has very little influence on other lexical-processing tasks, such as naming or lexical decision; for example, see Andrews, 1992; Treiman, Mullennix, Bijeljac-Babic, & Richmond-Welty, 1995.) One question that arises from this apparent lack of an influence of bigram frequency is why there are influences of neighbors only when the neighbors share more than two letters.

In addition to bigram frequency, one might ask whether positional frequency influences letter recognition. Positional frequency refers to the probability that a given letter(s) will occur in a given position within a word. Mayzner and Tresselt (1965) tabulated the summed positional frequency for single letters, bigrams, trigrams, tetragrams, and pentagrams (Mayzner, Tresselt, & Wolin, 1965a, 1965b, 1965c) across a set of 20,000 words. This metric should reflect the orthographic structure across words within a given language. In fact, one might expect influences of such a metric to fall quite nicely out of the McClelland and Rumelhart-type model. In fact, Massaro, Venezky, and Taylor (1979) reported evidence of a large impact of summed positional frequency within a Reicher-type paradigm. Their results indicated that both summed positional frequency and a rule-based metric of orthographic regularity (see discussion below) were found to influence letter recognition performance. Thus, at least at the level of letter recognition, there does appear to be an influence of positional letter frequency in a Reicher-type paradigm. Because letter position must be coded in the McClelland and Rumelhart model, one might expect this effect to naturally fall from the combined facilitatory and inhibitory influences across lexical-level representations. However, there are some limitations to such harsh coding. As discussed below in the section on orthographic neighborhood effects, the coding of position of letters within words has become a very active area of research recently.

In sum, the interactive activation model provides a cogent quantitative account of what appears to be evidence of multiple levels within the processing system working in concert to influence letter recognition (for an alternative view, see Massaro & Cohen, 1994). A particularly important aspect of this model is that "other" similar lexical-level representations appear to have an influence on the ease of recognizing a given letter within a word. It appears that letter- or word-level representations do not passively accumulate information, as in a logogen-type model (see Morton, 1969), but letters and words appear to be recognized in the context of similar representations that either reinforce or diminish the activation at a given representation. We shall now turn to some discussion of the dimensions that define "similarity" in such networks.

B. Getting from Letters to Words: Influences of Sublexical Levels of Organization

The journey from letters to words has been a central concern in word recognition models. Although there are many distinct issues that arise in this area, one of the major theoretical issues has been the specification of the "rules" that are used in translating an orthographic pattern into an acceptable lexical/phonological representation. Unfortunately, as we shall see, such a translation process is far from easy in the English orthography.

1. Specifying the "Rules" of Translation

One of the most evasive goals encountered in the analysis of English orthography is the specification of the functional unit(s) of sub lexical organization. An obvious spelling-to sound mapping might involve a simple one-to-one correspondence between graph emic units (single letters or letter clusters) and phonemes. Obviously, such an analysis fails relatively quickly in English because some graphemes, like PH, can serve as one phoneme in words like PHILOSOPHY, and two phonemes in a word like UPHILL. Likewise, even single letters are quite ambiguous such as the C in the word CAT and CIDER. English orthography simply does not allow a one-to-one mapping of spelling to sound.

One of the most evasive goals encountered in the analysis of English orthography is the specification of the functional unit(s) of sub lexical organization. An obvious spelling-to sound mapping might involve a simple one-to-one correspondence between graphemic units (single letters or letter clusters) and phonemes. Obviously, such an analysis fails relatively quickly in English because some graphemes, like PH, can serve as one phoneme in words like PHILOSOPHY, and two phonemes in a word like UPHILL. Likewise, even single letters are quite ambiguous such as the C in the word CAT and CIDER. English orthography simply does not allow a one-to-one mapping of spelling to sound.

Although a simple mapping of spelling to sound may not work for all words, it is still possible that one may gain considerable insight into the vast majority of words via an analysis of the regularities in the orthography. Such an enterprise was undertaken in a number of large-scale studies of English orthography in the late 1960s and early 1970s (e.g., Haas, 1970; Hanna, Hanna, Hodges, & Rudorf, 1966; Venezky, 1970; Wijk, 1966). For example, Venezky coded the grapheme-to-phoneme correspondences across a set of 20,000 medium- to high-frequency words. Through an in-depth analysis of the consistency of grapheme-to-phoneme patterns, Venezky distinguished between two large classes of grapheme-to-phoneme correspondences. Predictable patterns are those which can be based upon the regular graphemic, morphemic (minimal meaningful units, e.g., REDISTRIBUTION RE DISTRIBUTE TION), or phonemic features of the words in which they occur, whereas, unpredictable patterns do not appear to fit within any predictable class (e.g., CHAMOIS). The important question is to what degree are patterns predictable when one considers similarities across words within the language. For example, some correspondences appear to be relatively invariant (predictable invariant patterns), e.g., the grapheme F always corresponds to the sound /f/ with the only exception being in the word OF. On the other hand, other graphemes have many variations, each of which appear to be relatively predictable (predictable variant patterns). For example, the letter C most typically corresponds to the phoneme /K/, but corresponds to the phoneme /S/ in many words when it is succeeded by the letter I, Y, or E.

As Henderson (1982) points out, there are a number of sub lexical constraints within the grapheme-to-phoneme system in English, which are called phono tactic constraints. For example, because certain stop consonant sequences are not permissible in English (e.g., /b/p/ and /p/b/), whenever one is confronted with such a sequence of letters (e.g., PB or BP) the correspondence is such that the first phoneme is silent (e.g., SUBPOENA). Thus, in this case, the phonological constraints of the language drive the grapheme-to phoneme conversion of the spelling patterns. There also appear to be predictable constraints on the grapheme-to-phoneme mapping that are derived at the morphemic and syllabic levels. For example, the graph emic sequence MB corresponds to two separate phonemes when it segments syllables such as in ambulance and amber, but only one phoneme at word ending positions, such as in tomb and bomb. Unfortunately, as Henderson points out, the situation becomes somewhat more complex when one considers that MB also only corresponds to one phoneme when it precedes inflectional affixes (e.g., bombing), but not when it precedes other morphemes (bombard). Moreover, there appear to be other rule-type constraints that are simply based upon allowable grapheme to-phoneme correspondences in particular positions within words. For example, the CK spelling pattern corresponds to the phoneme /K/, but the CK pattern does not occur at the beginning of words; in these later cases, the C to /K/ correspondence or the K to /K/ correspondence occurs. Using sophisticated permutation analyses, Kessler and Treiman (2001) have also shown that the spelling-to-sound consistency of a syllabic segment (i.e., onset, vowel, and coda) increases substantially when the other two segments are taken into account. These results may support the contention that English spelling is not as chaotic or irregular as popularly thought (Kessler & Treiman, 2003).

For demonstrative purposes, we have only touched upon some of the problems that one encounters in attempting to understand the regularity of spelling-to-sound correspondences in English orthography. Although ultimately it may be possible to specify such grapheme-to-phoneme rules in English, it is noteworthy that even with the relatively complex rule system developed by Venezky, and others, Coltheart (1978) estimated that 10–15% of the words would still be unpredictable, i.e., irregular. Likewise, Wijk (1969) notes that about 10% of the words will not fit his Regularized Inglish. This may be an underestimate, because as Henderson points out, of the most common 3000 words, as many as 21% violate Wijk's regularization rules. Interestingly, Coltheart, Curtis, Atkins, and Haller (1993), using a learning algorithm to generate grapheme-to-phoneme rules, found that these rules mispronounced 22% of monosyllabic words, a figure which is consistent with Henderson's estimate. Also, because of computational limits inherent in two-layer networks (Hinton & Shallice, 1991), the two-layer network of Zorzi, Houghton, and Butterworth (1998) dual-process model was found to be incapable of learning exception words, and it failed to learn correct phonological codes for about 19% of its 2774 monosyllabic word corpus.

Of course, even if one could develop a rule-based system of spelling-to-sound translation that would accommodate all words in English, this would not necessarily indicate that such a rule-based system is represented in readers of English. In fact, even if such a rule-based system were represented, this would not be sufficient evidence to indicate that such rules are critical in fluent word recognition. Hence, instead of providing a detailed discussion of the enormously complex rule systems that have been developed to capture the mapping of orthography onto phonology in English, the present discussion will focus on the empirical evidence regarding how readers use sub lexical information in word recognition tasks. The interested reader is referred to Henderson (1982), Wijk (1966, 1969), and Venezky (1970) for excellent treatments of the search for rule-based translations of spelling-to-sound in English (for a description of the algorithms used to identify single-letter, multilateral, and context-sensitive rules, see Coltheart et al., 1993).

2. If Not Rules, Then What? The Controversy Regarding Dual-Route and Single-Route Models of Pronunciation

1) Dual Route Perspective

If it is unlikely that there will be a limited number of rules that specify the translation from spelling-to-sound in English (i.e., an assembled route), it is possible that there is a second route (the lexical or direct route) that also plays a role in recognizing words. In the second, lexical, route the reader may map the orthographic string onto a lexical representation and then access the programs necessary for pronouncing a given word aloud either directly from that representation or via access to a semantic representation. Figure 2 displays the dual-route cascaded (DRC) model of word reading developed by Coltheart, Rastle, Perry, Langdon, and Ziegler (2001). In their model, the lexical route is a straightforward extension to the interactive-activation model discussed above. One notable difference is that in the Coltheart et al. model, separate lexicons exist for orthography and phonology.

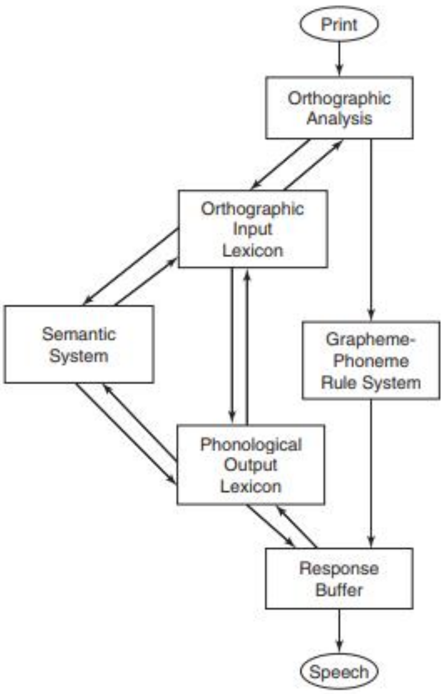


Figure 2. Coltheart et al.'s (2001) DRC model of word recognition.

It is important to note here that because the world's orthographies differ with respect to the regularity of spelling-to-sound correspondences, orthographies also appear to differ with respect to the weight placed on the assembled and lexical routes. For example, if the alphabetic system in a given language is unequivocal in mapping orthography to phonology, as in a language such as Serbo-Croatian, then one might find little or no impact of the lexical route in speeded pronunciation performance (Frost, Katz, & Bentin, 1987). The reader can rely totally on the assembled route, because it always produces the correct response. However, in English, and even to a greater extent in other languages such as in Hebrew (e.g., Frost et al., 1987), the mapping between orthography and phonology is far less transparent. Hence, one should find increasing lexical effects in speeded pronunciation performance as one decreases the transparency of the spellingto-sound correspondences (also referred to as the orthographic depth hypothesis). In support of this prediction, Frost et al. have reported larger frequency and lexicality effects in Hebrew compared to English which in turn produced larger effects compared to Serbo-Croatian. Similarly, there is evidence that readers of a shallow orthography like Serbo-Croatian make lexical decisions based on a prelexically computed phonological code; in contrast, phonological effects are relatively difficult to obtain in English lexical decision (Frost, 1998). Thus, comparisons across orthographies that differ with respect to the regularity of the spelling-to sound correspondence support the notion that two routes are more likely in languages that have relatively deep orthographies.

If the inadequacy of a rule-based system demands a lexical route in English orthography, then one might ask what evidence there is for a role of an assembled route. Why would subjects ever use an assembled route to name a word aloud, if, by necessity, there must be a lexical route? One piece of evidence that researchers originally identified is the relative ease with which individuals can name nonwords (e.g., blark) aloud. Because nonwords do not have a direct lexical representation, it would appear that a nonlexical route is necessary for naming nonwords. However, this piece of evidence was soon disabled by evidence from activation-synthesis-type approaches (e.g., Glushko, 1979; Kay & Marcel, 1981; Marcel, 1980), in which the pronunciation of a nonword could be generated by the activation of similarly spelled words. Activation-synthesis theorists argued that pronunciation performance is always generated via analogies to words represented in the lexicon, thus minimizing an important role for the assembled route.

However, there is a second, and more powerful, line of support for the role of an assembled route in English that involves the performance of acquired dyslexics, who appeared to produce a double dissociation between the two routes. Specifically, one class of dyslexics, surface dyslexics, appears to have a selective breakdown in the lexical route, but have an intact assembled route. These individuals are likely to regularize irregular words and exception words, e.g., they might pronounce PINT such that it rhymes with HINT (e.g., Marshall & Newcombe, 1980; McCarthy & Warrington, 1986; Shallice, Warrington, & McCarthy, 1983). A second class of acquired dyslexics, phonological (deep) dyslexics, appears to have an intact lexical route but an impaired phonological route. These individuals can pronounce irregular words and other familiar words that have lexical representations, however, when presented a nonword that does not have a lexical representation there is considerable breakdown in performance (Patterson, 1982; Shallice & Warrington, 1980). The argument here is that phonological dyslexics have a selective breakdown in the assembled route. Recently, Coltheart et al. (2001) simulated these two acquired dyslexias in the DRC model by selectively lesioning different components of the model. Specifically, surface dyslexia was simulated by lesioning the orthographic lexicon, while phonological dyslexia was simulated by dramatically slowing the sublexical process. These simulations nicely mimicked the neuropsychological data. For example, the degree of impairment of the orthographic lexicon produced regularization error rates that correlated highly with actual regularization error rates exhibited by surface dyslexics of varying severity. Furthermore, the model also correctly simulated the pseudohomophone advantage shown by phonological dyslexics. Specifically, these individuals pronounce pseudohomophones (e.g., BRANE) more accurately than non-pseudohomophonic nonwords (e.g., BRONE), reflecting the larger impact of the lexical route as the influence of the sublexical route is decreased (for a review of pseudohomophone effects in naming performance, see Reynolds & Besner, 2005a).

2) Parallel distributed processing

Although it would appear that there is compelling evidence for a dual-route architecture, there are important alternative models that have been developed by Seidenberg and McClelland (1989) and Plaut, McClelland, Seidenberg, and Patterson (1996) that also do an excellent job of handling some of the major findings that were originally viewed as strong support for the dual-route model. These parallel-distributed-processing (PDP) models could be viewed as a second generation of the original McClelland and Rumelhart (1981) model of letter recognition described above. One of the major differences between the two classes of models is that the later models were specifically developed to account for lexical tasks such as word pronunciation and the lexical decision task, whereas, the McClelland and Rumelhart model was developed in large part to account for letter recognition performance. A second major difference between the models is that the McClelland and Rumelhart model involves localized representations for the major processing codes (i.e., features, letters, and words), whereas, the later models involve distributed representations, e.g., there is not a single representation that reflects the word DOG. A third difference is that the McClelland and Rumelhart model assumes the existence of a specific architecture (i.e., sets of features, letters, and words along with the necessary connections), whereas, the latter models attempts to capture the development of the lexical processing system via the influence of a training regime. However, given these differences, both models account for performance by assuming a flow of activation across a set of relatively simple processing units and have been detailed sufficiently to allow for mathematical tractability. We shall now turn to a brief introduction to the Seidenberg and McClelland model, which was the first in a series of parallel distributed processing models of word recognition.

As shown in Figure 3, the Seidenberg and McClelland model involves a set of input units that code the orthography of the stimulus and a set of output units that represent the phonology entailed in pronunciation. All of the input units are connected to a set of hidden units (units whose only inputs and outputs are within the system being modeled, i.e., no direct contact to external systems, see McClelland & Rumelhart, 1986, p. 48), and all of the hidden units are connected to a set of output units. The weights in the connections between the input and hidden units and the weights in the connections between the hidden units and phonological units do not involve any organized mapping before training begins. During training, the model is presented an orthographic string which produces some phonological output. The weights connecting the input and output units are adjusted according to the back-propagation rule, such that the weights are adjusted to reduce the difference between the correct pronunciation and the model's output. During training, Seidenberg and McClelland presented the model with 2897 English monosyllabic words (including 13 homographs, resulting in 2884 unique letter strings) at a rate that is proportional to their natural frequency of occurrence in English. The exciting result of this endeavor is that the model does a rather good job of producing the phonology that corresponds to regular words, high-frequency exception words, and even some nonwords that were never presented. Although there is clearly some controversy regarding the degree to which the model actually captures aspects of the data (e.g., see Besner, 1990; Besner, Twilley, McCann, & Seergobin, 1990), the fact that it provides a quantitative account of aspects of simple pronunciation performance (without either explicit Venezky-type rules or even a lexicon) is quite intriguing and it presented a powerful challenge to the available word-recognition models.

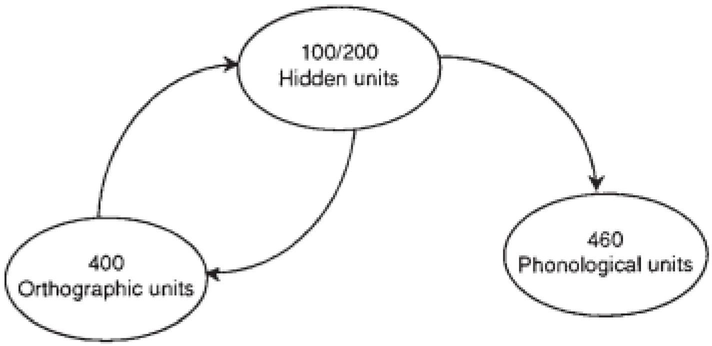


Figure 3. Seidenberg and McClelland's (1989) implemented connectionist architecture.

One of the more important results of the Seidenberg and McClelland model is its ability to capture the frequency by regularity interaction. This finding was initially viewed as rather strong support for a dual-route model (cf., Andrews, 1982; Monsell, Patterson, Graham, Hughes, & Milroy, 1992; Paap & Noel, 1991; Seidenberg, Waters, Barnes, & Tanenhaus, 1984a). The interaction is as follows: for high-frequency words, there is very little impact of the correspondence between orthography and phonology (but see Jared, 1997), whereas, for low-frequency words there is a relatively large impact of such a correspondence. The dual-route framework accommodated this finding by assuming that for high-frequency words the frequency modulated lexical route is faster than the frequency independent assembled route, and hence, any inconsistent information from the assembled route does not arrive in time to compete with the pronunciation that is derived from the lexical route. For example, the incorrect assembled pronunciation for the high-frequency word HAVE (such that it rhymes with GAVE) should not arrive in time to compete with the fast and correct lexical pronunciation. However, if one slows up the lexical route by presenting a low-frequency word (e.g., PINT), then one finds that the assembled output has time to interfere with the lexically mediated route and hence response latency is slowed down. The important point for the dual-route model is that the output of a low-frequency lexically mediated response can be inhibited by the availability of phonological information that is produced via the assembled route.

Although the dual-route model provides a natural account for this interaction, this pattern also nicely falls from the Seidenberg and McClelland single route model. That is, the error scores produced by the model (a metric that is assumed to map onto response latencies) for high-frequency regular words and exception words are quite comparable, however, for low-frequency words, the error scores are larger for exception words than for regular words. Thus, one does not have to assume separate routes (or even a lexicon) to handle the frequency by regularity interaction, because this pattern naturally falls from the correspondences between the frequency of a particular spelling-to-sound correspondence even in a relatively opaque alphabetic system such as English. The interaction between frequency and regularity for a specific set of words, and the predictions from Seidenberg and McClelland's model for this same set of words are displayed in Figure 4.

Interestingly, the spelling-sound consistency of a word's neighborhood also influences naming performance, and this neighborhood effect appears to produce an additional influence above and beyond the grapheme-to-phoneme regularity (Glushko, 1979; Jared, McRae, & Seidenberg, 1990). Consistency refers to the degree to which similarly spelled words are pronounced similarly. In particular, studies of consistency have focused on the rime (i.e., the vowel and subsequent consonants in a monosyllabic word). A word that shares both the orthographic rime and phonological rime with most or all of its neighbors is relatively consistent, whereas a word that shares the orthographic rime with its neighbors but has a different pronunciation than most of its neighbors is relatively inconsistent. Regular words that have many "friends" (e.g., spoon is consistent because of moon, noon, etc.) are named faster than regular words that have many "enemies" (e.g., spook is inconsistent because of book, took, etc.). Jared et al. (1990) provided evidence that there are consistency effects in pronunciation primarily under conditions when the neighbors that have consistent spelling patterns (i.e., friends) are higher in frequency than the neighbors that have inconsistent spelling patterns (i.e., enemies). Such neighborhood frequency effects would appear to fall quite nicely from the Seidenberg and McClelland (1989) model. Alternatively, a rule-based model might suggest that the consistency of the neighbors defines the rules of translation from orthography to phonology. However, because of the difficulties noted above in specifying such rules, it is appealing that the Seidenberg and McClelland model can capture such neighborhood effects, without the appeal to rules.

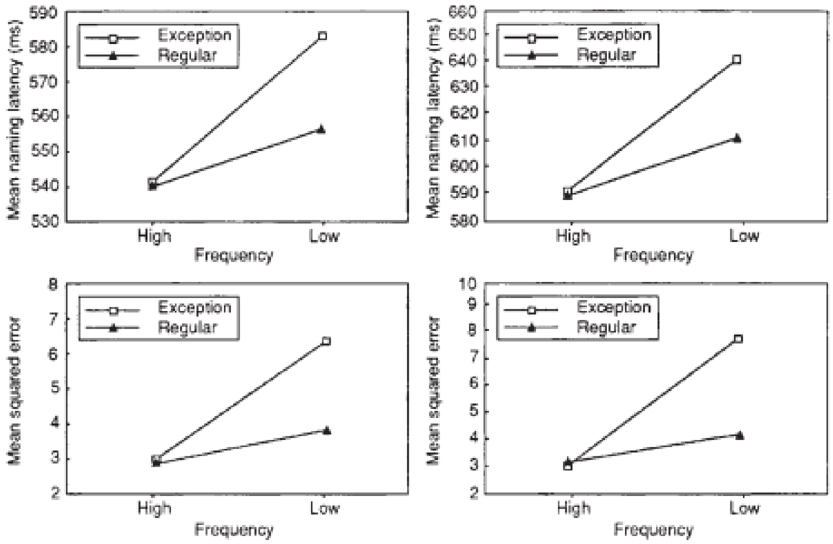


Figure 4. Results and simulations of the Seidenberg (1985, left graph) and Seidenberg et al. (1984a, Experiment 3, right graph) studies: Experimental results (upper graphs) and simulations from the Seidenberg and McClelland (1989) model (lower graphs).

3) Regularity vs. consistency revisited

Because many irregular words (i.e., words whose pronunciation violates grapheme phoneme correspondence (GPC) rules) are also inconsistent at the rime level, regularity and consistency have typically been confounded. However, these two dimensions are indeed separable (e.g., Andrews, 1982;

Kay & Bishop, 1987). Obviously, distinguishing regularity and consistency is important in testing contrasting predictions of models of word recognition. Specifically, the DRC model predicts large effects of regularity and small effects of consistency, and PDP models predict small effects of regularity and large effects of consistency. In general, the results of the studies that have distinguished between consistency and regularity have shown that rime consistency has a larger influence than regularity on latencies and errors than regularity (for discussion, see Cortese & Simpson, 2000; Jared, 2002). In fact, Cortese and Simpson found that the PDP model of Plaut et al. (1996) simulated the naming data on a selected set of words that crossed regularity and consistency better than the Coltheart et al. (2001) DRC model.

4) Regularity vs. consistency in words and no words

Of course, consistency is a continuous variable that can be measured at various levels (e.g., rimes, graphemes). In large-scale studies, Treiman and colleagues (Treiman, Kessler, & Bick, 2002; Treiman et al., 1995) have found that rime-level consistency is a better predictor of word naming performance than grapheme-to-phoneme level consistency. However, it appears that for no word naming performance, the pattern is a bit more complicated. For example, in contrast to the results by Treiman and colleagues regarding word naming performance, Andrews and Scarratt (1998) reported that nonword reading is affected more by consistency at the grapheme-to-phoneme level than by rimelevel consistency. Moreover, in their analysis of 20 nonwords (taken from Seidenberg, Plaut, Petersen, McClelland, & McRae, 1994) in which regularity and consistency pull in opposite directions, Cortese and Simpson (2000) found that grapheme-to-phoneme rules predicted the preferred pronunciation in 14 nonwords, whereas rime consistency predicted the preferred pronunciation in only 5 nonwords. Consider jind. The GPC rule for i is /I/, but consistency favors the /aInd/ pronunciation found in bind, blind, hind, mind, etc. Seidenberg et al. found that 23 of 24 participants pronounced jind in a fashion that is consistent with GPC rules. Therefore, it is quite possible that subjects may rely on different types of information when pronouncing a set of nonwords than when processing words.

Zevin and Seidenberg (2006) have recently claimed that consistency effects in nonword naming tasks are more consistent with the PDP perspective than the DRC perspective. By varying the training experience with each new run of a PDP model (also see Harm & Seidenberg, 1999), the model could be tested in terms of pronunciation variability (i.e., the degree to which pronunciations vary across subjects) that is exhibited by college readers (Andrews & Scarratt, 1998; Treiman et al., 2002). Both the PDP model and college readers exhibited considerable variability in their pronunciation of nonwords derived from inconsistent words (e.g., chead, moup), but not in their pronunciation of nonwords derived from consistent words (e.g., nust). This characteristic is difficult to assess in the DRC model because it is not clear how rules are acquired in the most recent model and how different versions of the model could be implemented. In addition to the insights provided regarding consistency effects in nonword naming, the extension of the models to individual variability, as opposed to overall mean performance, is an important next step in model development.

Zevin and Seidenberg (2006) have recently claimed that consistency effects in nonword naming tasks are more consistent with the PDP perspective than the DRC perspective. By varying the training experience with each new run of a PDP model (also see Harm & Seidenberg, 1999), the model could be tested in terms of pronunciation variability (i.e., the degree to which pronunciations vary across subjects) that is exhibited by college readers (Andrews & Scarratt, 1998; Treiman et al., 2002). Both the PDP model and college readers exhibited considerable variability in their pronunciation of nonwords derived from inconsistent words (e.g., chead, moup), but not in their pronunciation of nonwords derived from consistent words (e.g., nust). This characteristic is difficult to assess in the DRC model because it is not clear how rules are acquired in the most recent model and how different versions of the model could be implemented. In addition to the insights provided regarding consistency effects in nonword naming, the extension of the models to individual variability, as opposed to overall mean performance, is an important next step in model development.

5) Feedback consistency

Heretofore, we have been primarily discussing the directional feedforward mapping of orthography onto phonology in our consideration of regularity and consistency effects. For example, PINT is feedforward inconsistent because it does not rhyme with its orthographic neighbors (e.g., mint, hint, tint, etc.). However, there is another form of mapping which reflects a feedback influence. Specifically, feedback consistency reflects the manner in which a specific phonological pattern is spelled in different ways. Figure 5 illustrates the syllabic structure derived by linguistic distinctions between onsets and rimes (see further discussion below) and also shows how consistency can be computed along four dimensions: (a) feedforward onset, (b) feedforward rime, (c) feedback onset, and (d) feedback rime. For example, the rime in tone is feedback inconsistent because /on/ is spelled OWN as in GROWN, and OAN, as in MOAN. As one might guess, many words are inconsistent in both directions. Stone, Vanhoy, and Van Orden (1997) first decoupled feedforward consistency from feedback consistency and the effects of both variables were obtained in lexical decision performance.

In addition, reliable and equivalent feedback consistency effects were reported by Balota, Cortese, Sergent-Marshall, Spieler, and Yap (2004) for lexical decision and naming performance, whereas in French, Ziegler, Montant, and Jacobs (1997) found larger feedback consistency effects in lexical decision than in naming performance. The influence of feedback consistency in visual word recognition is theoretically important because it suggests that phonological activation provides feedback onto the orthographic representation (also see Pexman, Lupker, & Jared, 2001) during isolated visual word processing. However, it should also be noted that there is currently some debate regarding the unique effect of feedback consistency. For example, Peereman, Content, and Bonin (1998) have argued that feedback consistency effects in French are eliminated when familiarity is controlled (also see Kessler, Treiman, & Mullennix, 2005).

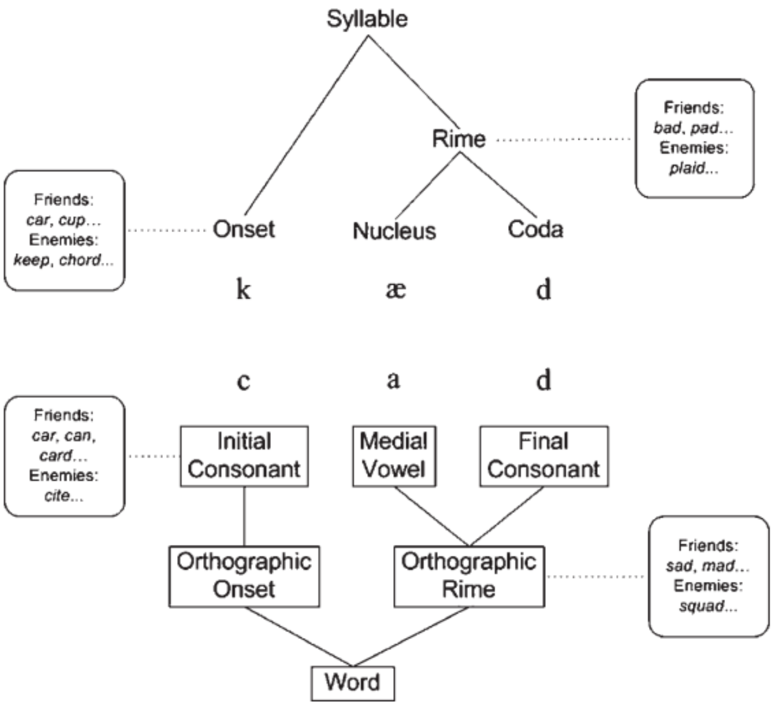


Figure 5. Feedforward and feedback onset and rime organization for single syllabic structure.

6) Potential problems with Seidenberg and McClelland model

Consistency effects would appear to arise naturally from the PDP architecture developed by Seidenberg and McClelland (1989). Although this model provided an interesting alternative to the dual-route model, it also generated a number of important problems that needed to be resolved (for a discussion of these issues, see Coltheart et al., 1993). First, it is unclear how such a model might handle the fact that some acquired dyslexics appear to only have an intact assembled route, while others appear to only have an intact lexical route (for some discussion of this issue, see Patterson, Seidenberg, & McClelland, 1989). Second, as described below, it appears that meaning-level representations can influence pronunciation and lexical decision performance. Thus, without some level of semantic input, it is unclear how an unembellished Seidenberg and McClelland model could account for such effects. Third, Besner (1990) and Besner et al. (1990b) have documented that the phonological error scores and the orthographic error scores do a rather poor job of simulating some characteristics of nonword performance. Fourth, the Seidenberg and McClelland model mapped error scores indirectly onto response latency instead of providing a direct metric for response latency.

7) Further developments in the PDP architecture

In response to the challenges to the Seidenberg and McClelland (1989) model, Plaut et al. (1996) substantively updated the representations and architecture of the PDP model (henceforth PMSP96) to address these problems. First, the model incorporated improved orthographic and phonological representations that allow it to not only correctly pronounce all the monosyllabic words in the training corpus, but also to name nonwords with a much greater facility. Second, in Seidenberg and McClelland's (1989) "triangle" model framework (see Figure 6), skilled reading is supported by the joint contributions of a phonological and a semantic pathway. Although only the phonological pathway was implemented in the Seidenberg and McClelland mode, a prototype semantic pathway was implemented in PSMP96 (see Simulation 4), which may be useful in accommodating meaning-level influences in word recognition (see Strain, Patterson, & Seidenberg, 1995). Third, the authors extensively discussed how the network could handle the acquired dyslexia data that, as described above, was central to the development of the DRC model. For example, it is possible to simulate phonological dyslexia, i.e., better word reading than nonword reading, by a selective impairment of the phonological pathway. Similarly, surface dyslexia, i.e., normal nonword but impaired exception word reading, was satisfactorily simulated after training a new network that incorporated an isolated, semantically supported phonological pathway. In normal readers, the semantic and phonological pathway work together to support the pronunciation of exception words. Should the semantic pathway be damaged, the semi-competent isolated phonological pathway manifests symptoms similar to that of surface dyslexia (Plaut, 1997). Of course, at this point one might ask whether the inclusion of a semantic "route" makes the PDP model functionally equivalent to a dual-route model. For example, does the network, over the course of training, partition itself into two sub-networks, one that handles regular words, and one that handles exception words? Plaut et al. (1996) tested this intriguing hypothesis and found little support for this contention.

Generally, the system did not fractionate itself such that one part learned spelling-sound rules and another part encoded the exceptions to these rules, but both components contributed to performance. Finally, it is important to note that the Plaut et al. model was a recurrent network that eventually settled into a steady state and hence response latencies could be evaluated in the model. That is, activation of phonological units changes over time as information in the system accumulates and is shared among network units. This property contrasts with the error measure that was used to evaluate the original Seidenberg and McClelland model. In the Plaut et al. model, when a word is recognized, its corresponding grapheme units become activated, and, in turn, this activation is propagated throughout the network.

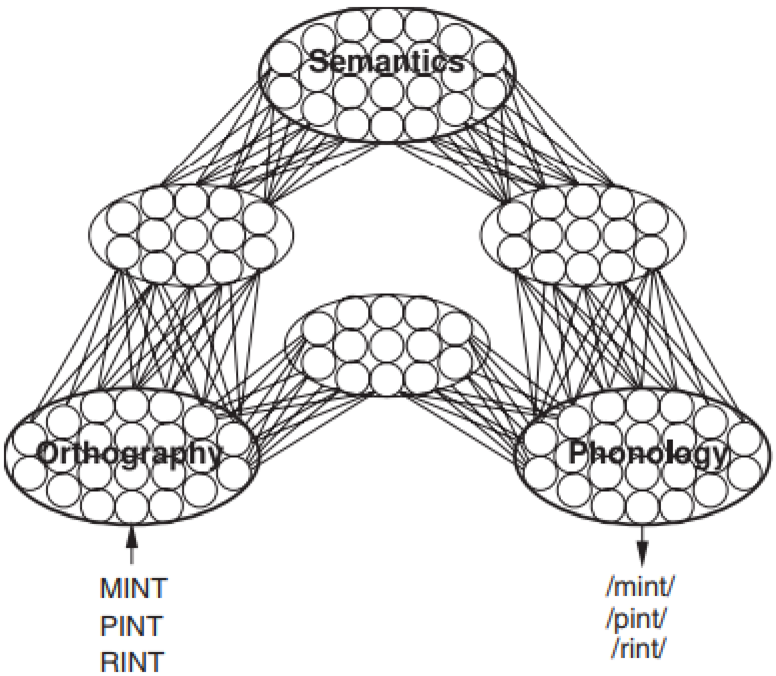


Figure 6. Seidenberg and McClelland's (1989) triangle connectionist framework for lexical processing.

Considerable debate continues between advocates of PDP and DRC approaches to word recognition. Although the PDP models seem ideally suited for handling consistency effects, the DRC model is particularly adept at handling data consistent with serial processing. Consider, for example, the position of irregularity effect. An irregular/inconsistent word can be irregular/inconsistent at the first phoneme position (e.g., chef), the second phoneme position (e.g., pint), the third phoneme position (e.g., plaid), or beyond (e.g., debris). Because the sub lexical process in the DRC operates in a serial fashion, it is more susceptible to earlier than later irregular/inconsistent sub lexical interference. In contrast, the PDP model processes words in parallel, and so does not predict a position of irregularity effect. Although there have been some methodological concerns noted (see Cortese, 1998; Zorzi, 2000), the evidence indicates that latencies are longer for words that have early irregular/inconsistent patterns than late irregular/inconsistent patterns (e.g., Coltheart & Rastle, 1994; Cortese, 1998; Rastle & Coltheart, 1999). These results, along with others (see Rastle & Coltheart, 2006), appear to support the serial component of the DRC-type framework over the parallel nature of the PDP framework. Of course, it is possible that such effects may ultimately reflect input or output processes beyond the scope of the currently implemented PDP framework. Indeed, Seidenberg (2005) acknowledges that important challenges exist, but the PDP approach accounts for many behavioral phenomena as well as providing a more natural interface between reading and underlying principles of the nervous system. In this light, Seidenberg argues that the PDP architecture is important because it generalizes well to other cognitive domains. However, a proponent of the DRC approach would argue that such generality should not outweigh the fact that the devil is in the details of the fit of a particular model of word recognition with the available evidence (see, for example, Rastle & Coltheart, 2006). Clearly, the debate continues.

More recent connectionist models of reading have shifted their emphasis from understanding how people pronounce letter strings aloud to understanding how meaning is computed (Seidenberg, 2005). For example, Harm and Seidenberg (2004) proposed a model that considers how the meaning of a word is computed by orthographic and phonological processes working cooperatively. It is also apparent that one glaring limitation of both dual-route and connectionist models is their inability to process multisyllabic words. One model that has made some progress in this respect is the connectionist multi trace memory model of Ans, Carbonnel, and Valdois (1998). While a full description of this interesting model is beyond the scope of this chapter, the Ans et al. model proposes two sequential procedures for reading: first, a holistic procedure that draws on knowledge about entire words, and if that fails, an analytic procedure that is dependent on the activation of subsyllabic segments. Including two reading procedures allows the model to name monosyllabic words, multisyllabic words, and nonwords, and also allows it to account for dissociations between skilled and pathological reading. Although the Ans et al.'s model may prima facie resemble the dual-route model, it does not compute phonology from orthography using different computational principles. Instead, pronunciation is always supported by the memory traces laid down by previously encountered exemplars.

3. Superletter Sublexical Codes : What's the Evidence for their Functional Role?

At this point, it should be noted that we have yet to discuss specific types of sublexical but supraletter influences on word recognition. We have generally grouped together a set of effects under the regularity/consistency umbrella, focusing on the theoretical implications of such effects for current models. We shall now turn to a brief discussion of three distinct levels of sublexical representation that have been at the center of this area of research: onsets and rimes, morphology, and syllables. The goal here is to simply acquaint the reader with the attempts that have been used to decompose the sublexical units.

1) Onsets and rimes

As noted earlier, researchers have made a distinction between the onset and rime unit within syllables. For example, Treiman and her colleagues (e.g., Treiman, 1989; Treiman & Chafetz, 1987; Treiman & Danis, 1988; Treiman & Zukowski, 1988) have argued that there is an intermediate level of representation in lexical processing between graphemes and syllables (also see Kay & Bishop, 1987; Patterson & Morton, 1985). They argue that syllables are not simply strings of phonemes but there is a level of subsyllabic organization that is used both in speech production and recognition of visual strings. This subsyllabic distinction is between the onset and rime of a syllable. The onset of a syllable can be identified as the initial consonant or consonant cluster in a word. For example, /s/ is the onset for sip, /sl/ is the onset for slip, and /str/ is the onset for strip. The rime of a word involves the following vowel and any subsequent consonants. For example, in SIP, SLIP, and STRIP, /Ip/ would be the rime. Thus, syllables have a subsyllabic organization in that each syllable is composed of an onset and a rime.

Although our primary interest is in visual word processing, it is interesting to note that there has been evidence from a number of quite varied research domains that supports the distinction between onsets and rimes in English. For example, there is evidence for this distribution from the types of speech errors that speakers produce (Dell, 1986; MacKay, 1972), the distributional characteristics of phonemes within syllables (Selkirk, 1982), along with the types of errors that subjects produce in short-term memory tasks (Treiman & Danis, 1988). Thus, the support for the onset and rime distinction clearly extends beyond the work in visual word recognition, and is driven more by phonological principles that have been developed in linguistics.

In one of the first studies addressing onset and rime organization in visual word recognition, Treiman and Chafetz (1987) presented strings like FL OST ANK TR to subjects with the task being to determine whether two of the strings in these four strings of letters could be combined to form a real word. In this case, one can see that FL and ANK can be combined to produce FLANK, with FL corresponding to the onset of the word FLANK and ANK corresponding to the rime. Now, consider performance in conditions where the strings again correspond to words but they are not broken at onsets and rimes. For example, a subject might be presented FLA ST NK TRO. For these items, the correct answer is again FLANK, but now the FLA and NK do not correspond to onsets and rimes. The results of the Treiman and Chafetz experiments indicated that anagram solutions were better when the breaks corresponded to onset-rime divisions compared to when the breaks did not. A similar pattern was found in a lexical decision task. In this study, the items were again presented such that there was either a break that matched the onset-rime division (e.g., CR//ISP, TH//ING) or a break that did not match the onset-rime division (e.g., CRI//SP and THI//NG). The results indicated that lexical decisions were reliably faster when the break matched the onset-rime division. Thus, Treiman and Chafetz argued that onset and rime units play a role in visual word recognition.

2) Syllables

If the distinction between onsets and rimes plays a functional role en route to word recognition then one would also expect a functional role for the syllable. At this level, it is quite surprising that there has been considerable disagreement regarding the role of the syllable in visual word recognition. For example, Spoehr and Smith (1973) argued for a central role of the syllable, whereas, Jared and Seidenberg (1990) have questioned the role of the syllable as a sublexical unit. In fact, as Seidenberg (1987) points out there is even some disagreement regarding where syllabic boundaries exist. For example, according to Howard's (1972) rules that emphasize intrasyllabic consonant strings surrounding a stressed vowel, CAMEL would be parsed as (CAM)(EL), whereas, according to Selkirk's (1980) more linguistically based view that emphasizes the maximal syllable onset principle CAMEL would be parsed (CA)(MEL). Obviously, before one can address the functional role of the syllable in visual word recognition, one must have some agreement on how to parse words into syllables. Fortunately, for the majority of words, there is agreement on how words are parsed into syllables.

The question here of course is whether a word like ANVIL is parsed into (AN)(VIL) en route to word recognition. It should again be emphasized here that the concern is not whether subjects have access to syllabic information, surely they must, i.e., most subjects can accurately decompose most words into syllables. The more important issue is whether this information is used in accessing the lexicon for visually presented words. Prinzmetal, Treiman, and Rho (1986) reported an intriguing set of experiments that investigated the impact of syllabic structure on early level perceptual operations in word recognition. These researchers used a paradigm developed by Treisman and Schmidt (1982) in which feature integration errors are used to examine perceptual groupings. The notion is that if a set of strings (e.g., letters or digits) forms a perceptual group then one should find migration of features (e.g., colors) toward that group. In the Prinzmetal et al. study, subjects were presented with words such as ANVIL and VODKA. At the beginning of each trial, subjects were given a target letter with the task being to report the color of the target letter that would appear in the upcoming display. After the target letter was designated, subjects were presented a letter string with each of the letters in different colors. The data of interest in such studies are the types of errors that subjects make as a function of syllabic structure. Consider the third letter position in the words ANVIL and VODKA. In the word ANVIL the third letter is part of the second syllable, whereas, in the case of VODKA the third letter is part of the first syllable. Now, if the syllable produces a perceptual grouping, then one might expect errors in reporting the colors such that the D in VODKA might be more likely to be reported in the same color of the O, compared to the K, whereas, the V in ANVIL might be more likely to be reported in the color of the I, compared to the N. This is precisely the pattern obtained in the Prinzmetal et al. study.

It is interesting to note here that Adams (1981) provided evidence that the letters that border adjacent syllables often have relatively low bigram frequencies. In fact, the NV and DK are the lowest bigram frequencies in the words ANVIL and VODKA. In general, if one considers relatively high-frequency bisyllabic words, there appears to be a decrease in frequency of the bigrams that occur at syllabic boundaries. This bigram trough may actually increase the likelihood of feature errors, due to the frequency of the orthographic neighbors of the target instead of an actual subsyllabic parsing en route to word recognition. Although Seidenberg (1987, Experiment 3) provided some initial evidence that the effects observed in the original Prinzmetal et al. paradigm were due to such bigram troughs, as opposed to actual syllabic boundaries, more recent work by Rapp (1992) found that one can obtain syllabic effects even when one controls for such bigram troughs.

The role of the syllable has not been implemented in most models of word recognition that have been primarily built to process monosyllabic words. One exception to this is the connectionist model proposed by Ans et al. (1998), discussed earlier. Based on the evidence discussed above and the findings from the literature on spoken word processing (e.g., Stevens & Blumstein, 1978), this model parses words into syllabic units in the phonological output. Presumably, this phonological output could serve as an access to a semantic system; however, this was not implemented in the current model.

More recently, Rastle and Coltheart (2000) have proposed a complex set of rules for syllable segmentation, stress assignment, and vowel reduction for disyllable words in their DRC model. In their study, the assignment of stress to a set of nonwords by the model was similar to that provided by human subjects. Also, words that violated the rules resulted in longer naming latencies, an effect that is consistent with predictions of the DRC model.

However, it is important to note that, like previous studies on regular and irregular monosyllabic words, regularity in the Rastle and Coltheart study was confounded with spelling-sound consistency (Chateau & Jared, 2003). In their naming study of disyllabic words, Chateau and Jared found that the feedforward consistency of the segment containing the first vowel grapheme and subsequent consonants and the second vowel grapheme predicted naming latencies and errors. Moreover, the consistency measures derived by Chateau and Jared nicely predicted the outcome reported by Rastle and Coltheart. Of course, if readers use sublexical rules when processing multisyllabic words, the DRC model would be better equipped to explain such a result, but consistency effects are better handled by PDP models. Clearly, more research on multisyllabic words is necessary to determine both the behavioral influence of syllables and stress patterns en route to word recognition and also the best way to model such effects.

3) Morphemes

Another sublexical unit that has received considerable attention in the literature is the morpheme. One of the most compelling reasons that morphemes might play a functional role in word recognition is the generative nature of language. Rapp (1992) provides CHUMMILY as an interesting example. Although we may have never encountered the nonword CHUMMILY, we may assume that it means something like in a chummy way or friendly because it appears to have the morphological form CHUMMY LY. Linguistic models of lexical representation assume that there is some base form of representation and a set of rules that are used to construct other forms of that item. The present question is whether a given form of a word such as JUMPED is parsed as (JUMP)(ED) en route to word recognition. As in the case of syllables, we are not questioning whether morphemes are represented in the processing system, the question is whether morphemic analyses play a role in processes tied to visual word recognition.

Much of the early theoretical and empirical work regarding the role of the morpheme in visual word recognition was originally developed by Taft and Forster (1975, 1976; also see Taft, 1979a, 1979b, 1985, 1987). They argued that readers first decompose polymorphemic words into constituent morphemes. Readers then access lexical files that are listed under the root morpheme. For example, if the word CHARACTERISTIC was presented, the reader would first access the root word CHARACTER and once this root word was accessed the subject would search through a list of polymorphemic words with the same root morpheme, e.g., Characteristic, Uncharacteristic, Characterized, Characteristically, Uncharacteristically, etc. There have been a number of studies reported in the literature that support the notion that there is a morphemic level of analysis in visual word recognition. For example, Taft (1979a, 1979b) found an effect of printed word frequency of the root morpheme (the sum of frequencies of all words with a given root) in lexical decision performance for items that were equated in surface frequencies (see, however, caveats by Bradley, 1979). This would appear to support the contention that root morphemes do play a special role in word recognition and it is not simply the raw frequency of the actual lexical string that is crucial.

Another approach to morphological analyses in word recognition involves long-term morphemic priming (e.g., Stanners, Neiser, & Painton, 1979a). In these studies, subjects are most often presented a sequence of lexical decision (word/nonword) trials. At varying lags within the sequence, subjects might be presented two forms of a given word with the same root. The interesting comparison is the influence of an earlier presentation of a given root form on later lexical decisions to the actual root. For example, if either JUMP or JUMPED is presented earlier in a lexical decision task, what impact does this presentation have on later lexical decision performance on the root form JUMP? Stanners, Neiser, Hernon, and Hall (1979b) found that both JUMP and JUMPED equally primed later lexical decisions to JUMP. Presumably, subjects had to access JUMP to recognize JUMPED and hence there was as much long-term priming from JUMPED as for the actual stem itself. Interestingly, Lima (1987) has found that mere letter overlap does not produce such an effect. For example, she reported that ARSON does not prime SON, but DISHONEST does prime HONEST. Thus, it does not appear that mere letter overlap is producing this long-term priming effect (for a summary of evidence favoring no orthographic accounts of morphemic priming effects, see review by Feldman & Andjelkovic, 1992).

Because the PDP perspective has achieved prominence as a general theory of language processing, research on morphological decomposition has taken on new theoretical significance. One main reason that this topic has received such attention is that distinct morphemic representations do not exist in PDP models (e.g., Plaut & Gonnerman, 2000; Rueckl, Mikolinski, Raveh, Miner, & Mars, 1997). Rather, morphemic effects are thought to emerge from interactions among orthography, phonology, and semantics (Gonnerman, Seidenberg, & Andersen, 2005). A recent cross-modal lexical decision study by Gonnerman et al. (2005) found support for this view. They reported that facilitation for visually presented targets was related to the semantic and phonological overlap found in prime–target pairs. In contrast, morphemic overlap did not produce additional facilitation above and beyond semantically and phonologically related items. For example, sneer facilitated snarl to the same degree as teacher facilitated teach. Also, weakly related pairs (e.g., lately-late) produced less facilitation than strongly related pairs.

Interestingly, Rastle, Davis, and New (2004) reported a morphological effect that was independent of semantics. In their lexical decision study, masked primes (presented for 42 ms) that maintained a morphological relationship only (e.g., corner-corn) facilitated targets as much as primes that maintained both a semantic and morphological relationship with the target (e.g., cleaner-clean), whereas a control condition (e.g., brothel, broth) did not produce priming. Thus, it appears from the Rastle et al. study that decomposition is somewhat independent of the semantic information available from the stem. This outcome seems more consistent with localist models (e.g., the DRC model) than distributed models (e.g., PDP models). However, given the Gonnerman et al. results discussed above, it is clear that further work is needed on this important topic.

We have only touched upon some of the very interesting issues that have arisen in morphological analyses in visual word recognition. We suspect that this will be an area of very active research in the future, and refer the reader to Baayen and Schreuder (2003), Feldman and Basnight-Brown (2005), and Sandra and Taft (1994) for more comprehensive treatments of this important area.

4. Lexical-Level Variables

By lexical-level variables, we refer to the impact of variables that have been quantified at the whole word level. For example, word frequency is a lexical variable. Specifically, a researcher can investigate the influence of the printed frequency of a given word (e.g., DOG vs. SILO) on word recognition task performance.

1) Length

One might ask whether there is a word length effect in visual word recognition tasks, as measured by the total number of letters in a given word. Obviously, if the letter is a crucial player in word recognition then one should find consistent effects of letter length. Interestingly, there has been some disagreement on this simple topic. There is clear evidence that longer words take more time in perceptual identification (McGinnies, Comer, & Lacey, 1952), and produce longer fixation durations in reading (see Just & Carpenter, 1980), but the effect of length in lexical decision and naming performance has been a bit more inconsistent (for a review, see New, Ferrand, Pallier, & Brysbaert, 2006).

The role of letter length in naming performance has been the focus of a number of recent studies. For example, Gold et al. (2005) found that individuals with a loss of semantic/lexical input, produced exaggerated length effects, compared to individuals with dementia of the Alzheimer's type. Gold et al. suggested that these results may be supportive of greater reliance on the serial sublexical route in individuals with semantic dementia. Consistent with this possibility, Weekes (1997) found length effects for nonwords and no length effects for words. Coltheart et al. (2001) interpreted the Weekes results as being critical to the DRC, i.e., the small or non-existent length effects for words is due to the parallel pathway used in the lexical route, whereas, the large length effects for nonwords reflects the serial analysis demanded by the sublexical route. In a study of speeded naming performance of over 2400 single syllable words, Balota et al. (2004) obtained clear effects of length that were modulated by word frequency. Moreover, lowfrequency words produced larger length effects than high-frequency words.

There is some controversy regarding length effects in the lexical decision task. Because the lexical decision task has been taken as a premier task to develop word recognition models, this is a troublesome finding (for a review, see Henderson, 1982). Chumbley and Balota (1984) reported relatively large length effects in the lexical decision task when the word and nonwords were equated on length and regularity. It is possible that inconsistent results with respect to past word-length studies using the lexical decision task may have been due to using a relatively small range of lengths of words. In this light, the recent study by New et al. (2006) is noteworthy. Specifically, they analyzed length effects in a dataset of lexical decision latencies to 33,006 words taken from Balota et al. (2002). They found an interesting quadratic relationship between length and lexical decision performance, such that there was a facilitatory effect from 3 to 5 letter in length null effect for 5–8 letters in length and a clear inhibitory effect for 8–13 letter words. The long words appear to demand some serial processing. Interestingly, the short words indicate that there may be an ideal length, based on the average length of words, and that very short words actually may produce a decrement in performance. Finally, it should also be noted that frequency does appear to modulate the length effect, since Balota et al. (2004) reported that length effects were larger in lexical decisions for low- than high-frequency words, similar to the pattern obtained in speeded naming performance mentioned above. Thus, the effects of word length in lexical decision performance appear to depend on both the frequency and the particular lengths of the words.

2) Word Frequency

The frequency with which a word appears in print has an influence on virtually all word recognition tasks. For example, word frequency effects have been found in lexical decision performance (e.g., Forster & Chambers, 1973), naming performance (e.g., Balota & Chumbley, 1984), perceptual identification performance (e.g., Broadbent, 1967), and online reading measures such as fixation duration and gaze duration measures (e.g., Rayner & Duffy, 1986; Schilling, Rayner, & Chumbley, 1998). This, of course, should not be surprising because printed word-frequency should be related to the number of times one experiences a given word; experience with an operation should influence the ease of performing that operation.

Although it would appear to be obvious why word-frequency modulates performance in word recognition tasks, the theoretical interpretations of such effects have been quite varied. For example, the activation class of models based in large part on Morton’s (1969, 1970) classic Logogen model, assume that frequency is coded via the activation thresholds in word recognition devices (logogens). High-frequency words, because of the increased likelihood of experience, will have lower activation thresholds than lowfrequency words. Therefore, in order to surpass a word recognition threshold, the activation within such a logogen will need to be boosted by less stimulus information for high-frequency words than for low-frequency words. Coltheart et al.'s (2001) DRC model nicely captures frequency effects via the activation patterns in the lexical route. The PDP models of Seidenberg and McClelland (1989) and Plaut et al. (1996) assume that frequency is coded in the weights associated with the connections between the units. Interestingly, there are hybrid models (e.g., Zorzi et al., 1998), which implement lexical and sublexical processing using connectionist principles, and so frequency effects could arise in both pathways.

A third class of word recognition models that we have yet to describe are referred to as ordered search models (e.g., Forster, 1976, 1979; Rubenstein, Garfield, & Millikan, 1970). According to these models, the lexicon is serially searched with high-frequency words being searched before low-frequency words. For example, as shown in Figure 7, Forster (1976) has argued that the lexicon may be searched via several indexing systems: orthographic, phonological, and syntactic/semantic access bins. Each of these bins involves a frequency ordered search, i.e., high-frequency words are searched before low-frequency words, and once the target is located the subject has immediate access to the word's master lexicon representation. Although such a model may seem cumbersome, Murray and Forster (2004) have recently provided intriguing evidence supporting this position, since rank frequency (as in rank in the search bin) appears to be a better predictor of word-frequency effects than actual log frequency values. It is noteworthy that there are additional models that are hybrids of the activation and search models such as in the Becker (1980), Paap et al. (1982), and the Taft and Hambly (1986) models. For example, Becker suggests that activation processes define both sensorily and semantically defined search sets. These search sets are then compared to the target stimulus via a frequency-ordered search process.

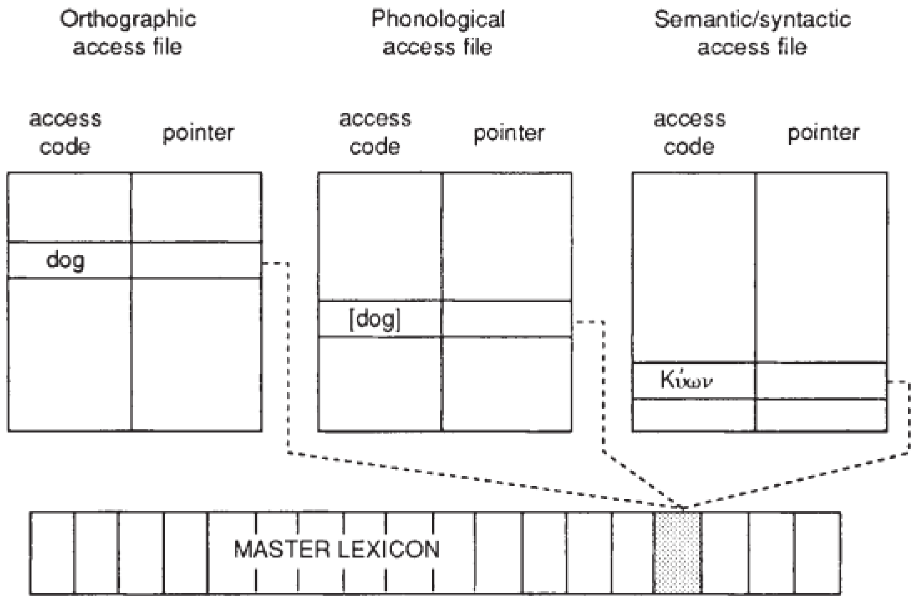


Figure 7. Architecture of Forster's (1976) serial search model of word recognition.

An important question that has arisen regarding word frequency effects is the locus of the effect in the tasks used to build models of word recognition. The models mentioned above all suggest that frequency is central to the interworkings of the models, as it should be. However, there is also evidence that suggests there are (a) decision components of the lexical decision task (Balota & Chumbley, 1984; Besner, Davelaar, Alcott, & Parry, 1984; Besner & McCann, 1987), (b) post-access components related to the generation and output of the phonological code in the pronunciation task (Balota & Chumbley, 1985; Connine, Mullennix, Shernoff, & Yelens, 1990), and (c) sophisticated guessing aspects of the threshold identification task (Catlin, 1969, 1973) that are likely to exaggerate the influence of word frequency. Because of the importance of task analyses, we will use this as an opportunity to review some of these issues regarding the lexical decision task.

Consider, for example, the Balota and Chumbley (1984) model of the lexical decision task displayed in Figure 8. Balota and Chumbley have suggested that because of the demands of the task, subjects place particular emphasis on two pieces of information that are obvious discriminators between words and nonwords, i.e., the familiarity and meaningfulness (FM dimension) of the stimuli. Nonwords are less familiar and also less meaningful than words. However, both words and nonwords vary on these dimensions; in fact the distributions may overlap (e.g., the nonword CHUMMINGLY is probably more familiar and meaningful than the low-frequency word TARADIDDLE). Frequency effects in the lexical decision task may be exaggerated because low-frequency words are more similar to the nonwords on the FM dimension than are high-frequency words. Hence, when there is insufficient information to make a fast "word" response the subject is required to engage in an extra checking process (possibly checking the spelling of the word). This time-consuming extra checking process is more likely to occur for low-frequency words than for high-frequency words, thereby exaggerating any obtained influence of word frequency. Hence, one should expect a larger influence of word-frequency in the lexical decision task than in the naming task, and in general this is what is found (see Balota et al., 2004). Balota and Spieler (1999) have implemented a hybrid model of the lexical decision task that not only accommodates word-frequency effects and other effects, but also accounts for the reaction time distributional aspects of performance, i.e., the shape of the reaction time distribution. It is also important to note that Ratcliff, Gomez, and McKoon (2004) have also argued that decision processes tied to the lexical decision task are critical in understanding word frequency effects, along with other variables, and have nicely modeled such effects with a single-process diffusion model.

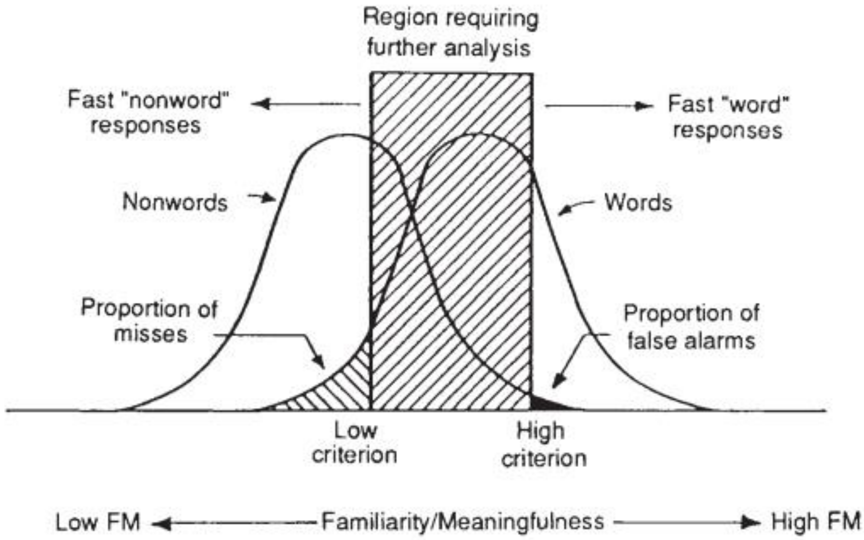


Figure 8. Balota and Chumbley's (1984) two-stage model of the lexical decision task.

There has been considerable controversy in the literature regarding the locus of word frequency effects in the tasks used to build word recognition models (e.g., see Andrews & Heathcote, 2001; Balota & Chumbley, 1990; Monsell, Doyle, & Haggard, 1989). Of course, the primary intent of the task analysis work is to caution researchers that not all word-frequency effects can be unequivocally attributed to access processes in the tasks that are used to measure word recognition. Although a full discussion of this work is beyond the scope of the present review, it is sufficient to note here that there is little disagreement that word-frequency influences processes involved in word recognition, and hence will need to be incorporated into all models of word recognition. However, as exemplified throughout this review, understanding the operations in the tasks used to build models of word recognition is a paramount first step in building adequate models.

3) Familiarity

A variable that is highly correlated with frequency is word familiarity. Familiarity is typically based on untimed ratings. For example, subjects may be asked to rate each word on a 7 point scale ranging from extremely unfamiliar to extremely familiar. The importance of familiarity norms was motivated by Gernsbacher (1984) who persuasively argued that the available printed word frequency norms by Kuera and Francis (1967) and Thorndike and Lorge (1944) may not be the most sensitive estimates of the impact of frequency of occurrence on lexical representations. For example, frequency norms typically do not take into account spoken word frequency, and are based on dated and relatively limited samples of word use. Gernsbacher (1984) pointed out that boxer, icing, and joker have the same objective frequency value (according to Kuera & Francis, 1967) as loire, gnome, and assay. Recently, there are a number of more extensive norms that have been developed based on a multifold increase in the sample size compared to the original norms (e.g., Baayen, Piepenbrock, & van Rijn, 1993; Burgess & Livesay, 1998; Zeno, Ivens, Millard, & Duvvuri, 1995). As one might expect, when comparing different frequency norms, the more recent norms are better predictors of both naming and lexical decision performance than the still commonly used Kuera and Francis (1967) norms (see Balota et al., 2004; Zevin & Seidenberg, 2002). Hopefully, cognitive science researchers who are investigating or controlling word frequency will begin to use these more recent norms.

Although the norms are becoming better, it is still the case that they are only a proxy for frequency of exposure. Hence, some researchers still argue that subjective familiarity ratings are a better measure of sheer exposure to a word. However, one might ask what sorts of information do subjects use when making an untimed familiarity rating? Standard instructions for familiarity ratings tend to be vague and may encourage the use of other types of information. For example, more meaningful stimuli tend to be rated more familiar. In fact, Balota, Pilotti, and Cortese (2001) found that the familiarity ratings of Toglia and Battig (1978) were related to meaningfulness, a semantic variable. As an alternative to standard familiarity ratings, Balota et al. (2001) had participant's rate monosyllabic words in terms of subjective frequency. Participants estimated how often they read, heard, wrote, said, or encountered each word based on the following scale: 1 never, 2 once a year, 3 once a month, 4 once a week, 5 every two days, 6 once a day, 7 several times a day. Balota et al. found that these ratings were less influenced by meaningfulness than the Toglia and Battig (1978) familiarity ratings. Hence, subjective frequency ratings may be more appropriate than traditional familiarity ratings because they are less influenced by semantic factors. Indeed, Balota et al. (2004) found that the subjective frequency ratings were highly predictive of both lexical decision and naming performance above and beyond a host of other correlated variables, such as objective word frequency, length, neighborhood size, spelling-to-sound consistency, etc.

4) Age of Acquisition

Within the past decade there has been considerable interest in the influence of the age at which words are acquired on various measures of lexical processing (for a recent review, see Juhasz, 2005). There have been a number of reports suggesting that age of acquisition (AoA) produces a unique influence on word recognition performance (e.g., Brown & Watson, 1987; Morrison & Ellis, 1995) above and beyond correlated variables such as word frequency. The intriguing argument here is that early acquired words could play a special role in laying down the initial orthographic, phonological, and/or semantic representations that the rest of the lexicon is built upon. Moreover, early acquired words will also have a much larger cumulative frequency of exposure across the lifetime.

There are at least two important methodological issues regarding AoA effects. The first concerns the extent to which AoA produces a unique effect in word recognition tasks like naming and lexical decision. One of the problems with assessing this issue is that AoA is correlated with many other variables, including length, frequency, and imageability. Moreover, one might expect an AoA effect not because early acquired words have a special influence on the lexicon, but rather because early acquired words have a greater cumulative frequency, even when objective frequency is held constant (for example, see Lewis, Gerhand, & Ellis, 2001). Although most studies have not teased these possibilities apart, Juhasz and Rayner (2003) found a unique effect of AoA in eye fixation data in reading and Bonin, Barry, Méot, and Chalard (2004) have demonstrated significant effects of objective AoA in word naming and lexical decision performance.

The second issue is concerned with whether or not AoA should be considered an outcome variable (Zevin & Seidenberg, 2002, 2004) or a standard independent (or predictor) variable. Zevin and Seidenberg have argued that AoA predicts word recognition performance because the age at which a word is learned is affected by many factors, and hence, this is related to the correlated variables issue noted above. They focus on frequency trajectory, which reflects the distribution of exposures that one has with words over time. Some words such as potty occur fairly frequently during early childhood but not adulthood, whereas other words such as fax occur frequently during adulthood, but not childhood. Therefore, frequency trajectory should influence AoA, and indeed the two variables are correlated. In addition, Zevin and Seidenberg (2004) examined the influence of frequency trajectory and cumulative frequency in naming. They found little evidence for frequency trajectory, whereas cumulative frequency produced a unique effect on naming performance (for an alternative interpretation of the Zevin & Seidenberg findings, however, see Juhasz, 2005). Given the potential theoretical importance of AoA, it appears that this variable will continue to be at the center of considerable empirical and theoretical work in the next several years.

5) Orthographic Neighborhood Effects

Although estimates vary, the average adult reader is likely to have about 50,000 words in their lexicon. Because these words are based on a limited number of 26 letters, there must be considerable overlap in spelling patterns across different words. One of the major tasks of an acceptable model of word recognition is to describe how the system selects the correct lexical representation among neighborhoods of highly related orthographic representations. Of course, it is possible that the number of similar spelling patterns may not influence lexical processing and that only a single representation must pass threshold for recognition to occur. However, as already mentioned, it appears that words are not recognized in isolation from other orthographically related representations.

Coltheart, Davelaar, Jonasson, and Besner (1977) introduced the orthographic neighborhood or N metric. N refers to the number of words that could be generated by changing only a single letter in each of the positions within a word. For example, the orthographic neighbors of the word FALL include MALL, FELL, FAIL, BALL, FULL, CALL, among others. There are two major ways that researchers have investigated the influence of N. First, consider the influence of the sheer number of orthographic neighbors. In naming performance, the results are rather straightforward: as the number of orthographic neighbors increases, response latency decreases, and this effect is larger for low-frequency words than high-frequency words (see Andrews, 1989, 1992; Balota et al., 2004). In contrast, in lexical decision performance, increases in N increase response latencies to nonwords, and for word targets the results range from facilitatory Andrews (1989, 1992; Forster & Shen, 1996) to no effect (Coltheart et al., 1977) to some conditions producing inhibitory effects (see, for example, Johnson & Pugh, 1994). In an excellent review of this literature, Andrews (1997) has argued that the variance across the studies of orthographic neighborhood size in lexical decision appears to be in part due to variability in list contexts (e.g., nonword type). It should also be noted that there is evidence of facilitatory effects of large Ns in semantic classification studies (Forster & Shen, 1996; Sears, Lupker, & Hino, 1999a). Finally, it should be noted that the evidence from eye-fixation patterns while people are reading indicate that there is an inhibitory effect of words with large Ns. Importantly, Pollatsek, Perea, and Binder (1999) have shown that with the same set of words that produces facilitatory effects in lexical decision performance, these words produce inhibitory effects in eye-fixation durations. Clearly, the effects of orthographic N are highly dependent upon the task constraints, and most likely a host of other variables such as individual processing speed (see, e.g., Balota et al., 2004).

A second way to investigate the influence of orthographic neighborhoods is to consider the frequency of the neighbors, i.e., does the stimulus have higher-frequency neighbors or lower-frequency neighbors? In lexical decision performance, there is evidence that targets with higher-frequency neighbors indeed produce inhibition in lexical decision performance, compared to words with lower-frequency neighbors (e.g., Grainger, 1990, 1992; Grainger & Jacobs, 1996; Carreiras, Perea, & Grainger, 1997, but see Pollatsek et al., 1999). However, there is even some conflict here, because in a series of experiments involving both naming and lexical decision performance, Sears, Hino, & Lupker (1995) found facilitation for low-frequency targets with large neighborhoods and higher-frequency neighbors.

Given that word recognition unfolds across time, it is not surprising that both frequency of the neighbors and the size of the neighborhoods should play a role in word recognition tasks. In this light, it is useful to mention the Luce and Pisoni (1989) neighborhood activation model, which they applied to auditory word recognition performance. This model takes into consideration target frequency, neighbor frequency, and neighborhood size via R. D. Luce's (1959) choice rule. Specifically, the probability of identifying a stimulus word is equal to the probability of the stimulus word divided by the probability of the word plus the combined probabilities of the neighbors. Of course, it is possible that the neighborhoods of the neighbors may play a role along with the degree of overlap of the neighbors. At this level, it is noteworthy that recent simulations by Sears, Hino, & Lupker (1999b) have shown that both the Plaut et al. (1996) and the Seidenberg and McClelland (1989) models appear to predict facilitatory effects of neighborhood size that are greater for low-frequency words than for high-frequency words, which is overall most consistent with the data in this area.

Facilitatory neighborhood effects for low-frequency words would appear to be difficult to accommodate within models that have a competitive interactive activation component (e.g., the DRC model of Coltheart et al., 2001, or the multiple read-out (MROM) model of Grainger & Jacobs, 1996). Specifically, the larger the neighborhood, the more competition one should find. Moreover, the facilitatory effects of N produce particular difficulties for serial search models, such as Forster's classic bin model. Specifically, the more items that need to be searched, the slower response latency should be. This is opposite to the most common pattern reported in this literature.

An interesting variation on the influence of orthographic N is the transposed letter effect. Specifically, Chambers (1979) and Andrews (1996) found that words like SLAT produce slower response latencies in lexical decision performance, because these items have a highly similar competitor SALT. Andrews (1996) also found this pattern in naming performance. Note that SLAT is not an orthographic neighbor of SALT, but is very similar because two letters in adjacent positions are switched. As Perea and Lupker (2003) have recently argued, the influence of transposed letter stimuli is inconsistent with most available models of word recognition, because these models typically code letters by positions within the words. These results are more consistent with recent input coding models such as SOLAR (Davis, 1999), and SERIOL (Whitney, 2001) that use spatial coding schemes for input of letters, that are not simply position specific (also see Davis & Bowers, 2004). Clearly, this is an important new area of research that extends the original work on orthographic N effects and has important ramifications of how the visual system codes the spatial position of the letters within words.

6) Phonological Neighborhood Effects

Although the influence of orthographic neighbors has dominated work in visual word recognition, it is quite possible that phonological neighbors may also play a role. Indeed, work by Yates, Locker, and Simpson (2004) has recently shown that lexical decision performance is facilitated by words with large phonological neighborhoods (also see Yates, 2005). Here, a phonological neighbor reflects a change in one phoneme, e.g., GATE has the neighbors HATE and GET, and BAIT. Yates et al. have also noted that previous studies of orthographic neighborhood size have typically confounded phonological neighborhood size. Although this is a relatively new area of exploration, it indeed is quite intriguing regarding the role of phonology in early access processes (see earlier discussion of feedback consistency effects), and has potentially important implications for how phonology is coded in the extant models (also see Ziegler & Perry, 1998).

5. Semantic Variables for Isolated Words

There have been a number of reports in the literature that indicate that semantic variables associated with lexical representations can modulate the ease of word recognition (see review by Balota, Ferraro, & Connor, 1991, of the early work in this area). This is an intriguing possibility because many models of word recognition would appear to indicate that the word must be recognized before the meaning of the word is determined. For example, within a logogen model, the lexical representation will need to reach threshold before the meaning of the word becomes available. How could it be otherwise? How could the system have access to the meaning without knowing what the stimulus is? Of course, this has some similarity to the word superiority effect described earlier wherein it was argued that the word level information is activated before the letters that make up the word have been recognized, via cascaded top-down activation. In fact, recent computational models by Coltheart et al. (2001) and Plaut et al. (1996) would appear to be able to handle such cascaded influences of meaning en route to making a speeded naming response.

Although there has been a considerable amount of work attempting to specify which semantic variables play a role in word recognition, much of this work has been open to alternative interpretations. Here, we shall briefly review this work emphasizing the primary findings with respect to each of the major variables.

1) Concreteness/Image ability Effects

Because concreteness is highly correlated with image ability, we will lump these variables together here. Concreteness refers to whether a word can be the object of a sense verb (e.g., touch, see, hear, etc.), whereas imitability typically involves subjects rating words on a low to high image ability scale. One might expect that high-image able words (e.g., CARROT) may be better recognized than low-image able words (e.g., FAITHFUL), because of the influence of the more salient referent being activated. Although the early evidence suggested that there were indeed effects of the concreteness/image ability variables (e.g., Boles, 1983; Day, 1977; Paivio & O'Neill, 1970; Rubenstein et al., 1970; Winnick & Kressel, 1965), some of this work was questioned because of the potential for confounding variables (see, for example, Schwanenflugel, Harnishfeger, & Stowe, 1988). However, there are indeed studies that are less susceptible to such criticism and have confirmed that there are concreteness/image ability effects in lexical decision, which are larger for low-frequency words than high-frequency words (e.g., de Groot, 1989; James, 1975; Kroll & Merves, 1986). Of course, this finding in and of itself is not terribly compelling evidence for an influence of meaning en route to word recognition performance, because one could argue that subjects place a premium on semantics in discriminating words from nonwords in the lexical decision task. Hence, the results from the naming task are indeed more noteworthy. Although the effects are clearly smaller, there is also evidence of an effect of concreteness/image ability in naming (e.g., Bleasdale, 1987). In the Balota et al. (2004) study of over 2000 monosyllabic words, they found that there was a reliable unique effect of image ability in naming (based on norms developed by Cortese & Fugett, 2004) after other variables were controlled, but this effect was quite small compared to lexical decision.

Image ability has played a special role in recent work exploring naming performance. Specifically, Strain et al. (1995) found an intriguing interaction between word frequency, spelling-to-sound consistency, and image ability.

They found that low-frequency words with inconsistent spelling to sound mappings produced the largest image ability effects. This was viewed as reflecting greater input from preexisting semantic representations for items with relatively low spelling to sound mapping, i.e., low-frequency inconsistent words, which they viewed as consistent with the tripartite connectionist framework, as exemplified by the Plaut et al. (1996) model. It should also be noted, however, that there is some controversy regarding potential correlated variables that may have contributed to this pattern (see Monaghan & Ellis, 2002; Strain, Patterson, & Seidenberg, 2002).

2) Meaningfulness

A second semantic variable that could play a role in word recognition is the meaningfulness of the stimulus. One way of measuring meaningfulness is simply to count the number of dictionary meanings for each word (for further discussion of different metrics of meaningfulness, see Millis & Button, 1989). Again, the early work in this area was controversial. For example, Jastrzembski (1981). Found initial evidence for a facilitatory effect of number of dictionary meanings, while, Gernsbacher (1984) argued that this was likely due to familiarity being confounded with meaningfulness. Azuma and Van Orden (1997) found an effect of number of meanings in lexical decision performance, but this seemed to depend on the relatedness of the meanings for a word. In fact, Azuma and Van Orden argued that the relatedness of the meanings is more important than the sheer number of meanings. As described below, this may be related to more recent notions of semantic connectivity. Balota et al. (2004) found a small and unique effect of subject rated meaningfulness that was larger in lexical decision than in naming performance. Finally, it is noteworthy that Rodd (2004) has recently provided evidence that the effect of number of meanings in speeded naming is larger for inconsistent spelling to sound mappings. This, of course, is consistent with the theoretically important observation of an increased influence of a semantic variable (image ability) for low-frequency inconsistent items, reported by Strain et al. (1995) described above. Likely due to familiarity being confounded with meaningfulness. Azuma and Van Orden (1997) found an effect of number of meanings in lexical decision performance, but this seemed to depend on the relatedness of the meanings for a word. In fact, Azuma and Van Orden argued that the relatedness of the meanings is more important than the sheer number of meanings. As described below, this may be related to more recent notions of semantic connectivity. Balota et al. (2004) found a small and unique effect of subject rated meaningfulness that was larger in lexical decision than in naming performance. Finally, it is noteworthy that Rodd (2004) has recently provided evidence that the effect of number of meanings in speeded naming is larger for inconsistent spelling to sound mappings. This, of course, is consistent with the theoretically important observation of an increased influence of a semantic variable (image ability) for low-frequency inconsistent items, reported by Strain et al. (1995) described above.

As noted, meaningfulness is typically defined by the number of dictionary meanings, which can vary in subtle but related ways. For example, the word DOG can mean the four legged animal, but it can serve as an adjective such as in "My car is a real dog," wherein the meaning of the word DOG is extended to another form. These might be considered different shades of the same meaning as opposed to distinct meanings of the word. In this light, there has also been some intriguing work investigating word recognition performance on homographs (e.g., the word ORGAN has two very different meanings referring to musical meaning and bodily system). It appears that such items can produce a facilitatory effect in both naming and lexical decision performance (see Hino & Lupker, 1996; Hino, Lupker, & Pexman, 2002). Interestingly, although one finds facilitation in naming and lexical decision, Hino et al. (2002) found inhibition in semantic categorization. These authors argued that only when attention is directed to retrieve semantic information, as in the semantic categorization task, will one find interference effects. A similar pattern was observed by Balota and Paul (1996) in a semantic relatedness judgment task. Finally, in neutral contexts, on-line measures of reading performance, as reflected by eye-fixation durations, suggest that there is interference when ambiguous words have relatively equally dominant interpretations (e.g., CLUB means to hit and organization, with similar frequencies), as if the meanings are competing for interpretation (for a review, see Duffy, Morris, & Rayner, 1988; Morris, this volume). Again, we find that task constraints strongly modulate the influence of a variable.

3) Grounding Semantics in Large-Scale Databases

There have been a number of recent attempts to ground semantics via analyses of large databases of natural language. This approach avoids some of the pitfalls in trying to quantify meaning as feature lists (e.g., the word DOG may include the features furry, barks, four-legged, pet) or some abstracted prototype (e.g., the modal DOG that is based on your experience with all DOGs). These more recent approaches include Burgess and Livesay's (1998) hyperspace analogue of language (HAL) and Landauer and Dumais' (1997) latent semantic analysis (LSA). HAL and LSA capture the meaning of words from the context in which a given word appears. Hence, the meaning of DOG is an evolving concept dependent upon an individual's experience with DOG in various linguistic contexts. Buchanan, Westbury, and Burgess (2001) have shown that estimates from HAL indeed predict lexical decision performance (for a detailed discussion of this work, see Burgess, this volume). It is noteworthy that an early study by Schwanenflugel, Harnishfeger, and Stowe (1988) provided evidence that a variable referred to as contextual availability can have an influence on isolated word recognition in lexical decision performance above and beyond influences of correlated variables such as concreteness, familiarity, length, etc. Contextual availability refers to how easily a subject is able to think of contexts in which a given word might occur.

An intriguing alternative approach has recently been developed by Steyvers and Tenenbaum (2005). They have utilized recently developed graph theoretic techniques to look at metrics of connectivity (along with other metrics) of meanings of words in a set of large-scale databases including Roget's (1911) Thesaurus, Miller's (1990) WordNet, and Nelson, McEvoy, and Schreiber's (1998) word association norms. Based on analyses of these databases, Steyvers and Tenenbaum have shown that semantic memory has a small-scale network structure in which a relatively small number of concepts serve as communication hubs for the rest of the semantic network. If semantic networks are represented in terms of the structure hypothesized by Steyvers and Tenenbaum, then words characterized by a high degree of connectivity with other words may be processed more quickly than words characterized by sparse connections. Indeed, Steyvers and Tenenbaum found evidence for such an effect in naming and lexical decision performance, above and beyond more standard lexical variables (also see Balota et al., 2004).

4) Additional Semantic Variables that Produce Effects in Isolated Word Recognition Paradigms

Because of space limitations, we shall only briefly mention a few other findings that would appear to indicate that meaning can have an early influence in word recognition performance. First, there is evidence that concreteness of a word can influence the time taken to generate an associate from that word (e.g., de Groot, 1989). Because subjects must recognize a word en route to generating an associate, this effect might be due to word recognition processes. Second, and along these same lines, Chumbley and Balota (1984) have found that the time taken to generate associates from one group of subjects can be used as a predictor of lexical decision performance for the same set of words when presented in isolation to a second group of subjects, above and beyond other related variables such as frequency, length, etc. Third, Whittlesea and Cantwell (1987) found that providing meaning for a nonword can produce a word-superiority effect, and also a study by Forster (1985) indicated that providing meaning for a nonword can produce a masked form priming effect in the lexical decision task. Both the word-superiority effect and the masked form priming effect would appear to tap relatively early lexical processes. Finally, there is evidence from masked semantic priming studies (reviewed below) suggesting that highly masked primes (that subjects apparently cannot consciously recognize) produce semantic priming effects, i.e., facilitate the processing of related targets compared to unrelated targets (see Holender, 1986, and the accompanying commentary for a discussion of the degree of conscious processing of the primes in these studies). At the very least, such threshold priming effects suggest that under presentation conditions that minimize conscious processing of the prime, meaning access can still occur.

5) Summary

The possibility that meaning-level representations play a role in isolated word recognition has relatively far reaching implications for current models of word recognition. Most of the available models emphasize the stages that subjects use in accessing the mental lexicon, with relatively little direct influence of meaning-level variables. However, when reminded that the role orthographic patterns play in reading is to convey meaning and not simply to convey lexicality then one might easily envisage an architecture that incorporates a relatively early influence of meaning. At this level, it should be no surprise that meaning-level representations may contribute to relatively early perceptual analyses and aid in constraining the percept, i.e., recognition of the word. Although recent connectionist and dual-route models of word processing acknowledge such effects, the devil is in the details of implementing such meaning-level influences.

6. Context/Priming Effects

Heretofore, we have primarily discussed the literature that deals with variables that influence isolated visual word recognition. Of course, readers typically encounter words in the context of other words. We now turn to a summary of the influences of contexts (hereafter referred to as primes) on word recognition processes. In these studies, two letter strings are typically presented and the researcher manipulates the relation between the two strings. For example, the strings may be orthographically related (COUCH-TOUCH), phonologically related (MUCH-TOUCH), semantically related (FEEL-TOUCH), or unrelated (NAIL-TOUCH). By manipulating the types of relationships between the primes and targets one can obtain evidence regarding the architecture of the word recognition system. For a more detailed discussion of this rich literature, see Neely (1991), Hutchison (2004), McNamara (2005) for important reviews of the semantic priming literature and Kinoshita and Lupker (2003b) for a volume dedicated to masked priming effects.

1) Orthographic Priming Effects

An interesting approach to identifying the access code in word recognition is the masked orthographic priming paradigm developed by Evett and Humphreys (1981, also see Humphreys, Besner, & Quinlan, 1988; Humphreys, Evett, Quinlan, & Besner, 1987). In this paradigm, subjects are briefly presented two letter strings that are both preceded and followed by pattern masks. The two letter strings vary in terms of orthographic, phonological, or semantic relatedness. Here, we focus on the orthographic priming conditions. There are a number of interesting findings in these masked priming studies: first, on most trials, subjects are unable to consciously identify the prime items and hence any influence of the prime items presumably reflects early access processes. Second, subjects are better at identifying the second letter string when it shares letters with the first letter string even though these shared letters are presented in different case. For example, relative to a baseline (e.g., harmless-ATTITUDE), there are priming effects for both identity priming (e.g., attitude-ATTITUDE) and form priming (e.g., aptitude-ATTITUDE). Third, in lexical decision, evidence for nonword repetition priming (e.g., flirp-FLIRP) is clearly less powerful than word repetition priming (Forster, 1998). Although earlier studies actually failed to find nonword repetition priming effects in lexical decision (see, for example, Forster and Davis, 1984), more recent studies have observed reliable effects (Bodner & Masson, 1997; Sereno, 1991). Fourth, in masked repetition priming studies, the effects of target word frequency and prime-target repetition are additive (Forster & Davis, 1984), a finding which is more consistent with search-class than with activationclass models of lexical access. Fifth, eye-tracking studies by Rayner, McConkie, and Zola (1980) using orthographic priming techniques have provided compelling evidence for a case independent orthographic code being used to access words in the parafovea while reading (for reviews, see Balota & Rayner, 1991; Rayner, 1998).

A particularly intriguing aspect of the masked priming literature is that within a range of short-duration primes, there is a relatively linear relationship between the duration of the masked prime and the magnitude of the priming effect (Forster & Davis, 1984). Specifically, a prime with a duration of 30ms produces a priming effect of about 30ms, whereas a prime with a duration of 20ms produces a priming effect of about 20ms. Forster (1998) has argued that this is most consistent with an Entry Opening process where the prime has the influence of opening the target's lexical representation, allowing the target to be processed more rapidly. This Entry Opening account of masked priming nicely accommodates the equivalent masked repetition effects for high-frequency and lowfrequency words, i.e., the masked prime has the effect of opening the lexical representation (Forster & Davis, 1984). However, it is unclear how the Entry Opening model accounts for nonword repetition priming effects, since nonwords, by definition, have no pre-existing lexical representations. To address such nonword effects, Bodner and Masson (1997) have proposed that masked priming effects are driven by a nonlexical locus, specifically, the retrieval of episodic memory traces established during previous encounters with the stimulus (for an episodic trace view of lexical processing, see Goldinger, 1998). This account implies that masked nonword primes operate nonlexically to facilitate orthographic processing (for an alternative explanation, see Forster, 1998).

Finally, task-specific effects have also been observed in masked priming. For example, there is evidence for a phenomenon called the masked onset priming effect. This effect was first reported by Forster and Davis (1991), who found that naming latencies to a target were facilitated when the prime and target shared the initial letter (e.g., save-SINK) compared to when they did not (e.g., farm-SINK). Further work by Kinoshita (2000) has revealed that this effect is position-dependent and is observed only when the initial onset (not the letter) is shared. For pairs bingo-BLISS, which has a common initial letter but different onsets (i.e., /B/ vs. /BL/), the effect was eliminated. Kinoshita argued that this supported a serial left-to-right procedure in naming performance, and may reflect articulatory planning rather than orthography-to-phonology computations (see also Schiller, 2004). The onset effect is only observed with tasks that require articulation, such as speeded naming, and not with lexical decision (Forster & Davis, 1991). Positing an articulatory nonlexical priming component for speeded naming may also explain why nonword repetition priming effects, which are equivocal in lexical decision, are more consistent in speeded naming (Masson & Isaak, 1999).

2) Phonological Priming Studies

There has been considerable debate concerning the role of phonological codes in word recognition (for an excellent review of this literature, see Frost, 1998). The extremes range from all words must be recognized via a phonological (assembled) code to the notion that many words (e.g., high-frequency words for skilled readers) are only accessed via an orthographic (addressed) code. Although there is controversy regarding the role of a phonological code in visual word recognition, there is considerably less debate regarding the importance of phonological codes in reading text, wherein, phonological codes produce representations that appear better suited for aspects of comprehension that place considerable demands on the working memory system (e.g., Baddeley, Eldridge, & Lewis, 1981; Besner, 1987; Slowiaczek & Clifton, 1980). It is possible that such phonological codes become active after lexical access has taken place in such reading studies. The more narrow issue here is whether phonological codes are necessary en route to word recognition. With this in mind, we now turn to the phonological priming literature.

Evett and Humphreys (1981) used the masked priming paradigm, described above, also to investigate the viability of a phonological access code, under conditions wherein conscious processing was limited. The results of this study indicated that there was priming for pairs that were orthographically and phonologically related (e.g., bribe-TRIBE) compared to pairs that were orthographically related but phonologically unrelated (break FREAK). Moreover, the effect occurred across case changes. In addition, in a similar masked priming paradigm, Humphreys, Evett, and Taylor (1982) found that identification accuracy was higher for targets (e.g., SHOOT) that followed homophonic primes (e.g., chute) compared to targets that followed graphemically related (e.g., short) or unrelated primes (trail). However, there was no facilitation from a nonword phonologically related prime (e.g., smorl-SMALL), suggesting a lexical locus for the priming effect.

Evidence for phonological mediation has also been obtained with an associative priming paradigm, which permits conscious, albeit brief, processing of primes. For example, Lukatela and Turvey (1994) compared priming effects across four conditions at different stimulus onset asynchronies (SOAs): standard semantic priming (e.g., TOAD-FROG), word homophonic priming (e.g., TOWED-FROG), nonword homophonic priming (e.g., TODE-FROG), and an orthographic control condition (e.g., TOLD-FROG). At short (i.e., 50ms) SOAs, the three related conditions produced comparable facilitation priming effects, compared to the control condition. However, at longer SOAs (i.e., 250ms), TODE became a stronger prime than TOWED. These findings reinforce the role of phonology in early visual lexical access, and also suggest that although word homophone primes (i.e., TOWED) are initially effective, they are quickly suppressed when the system detects the mismatch between their orthography and the addressed spelling of TOAD.

It is important to point out that the validity of the findings described above rests on the assumption that the orthographic control (e.g., TOLD) is as orthographically similar to the critical associate (e.g., TOAD) as the homophone (e.g., TOWED) (Pollatsek, Perea, & Carreiras, 2005). Some have failed to replicate the homophone/pseudohomophone advantage described above (see, for example, Davis, Castles, & Iakovidis, 1998) and Pollatsek et al. argued that this inconsistency may be due to imperfect matching of controls to homophones. After controlling for this potential confound, Pollatsek et al. still observed early phonological effects in a Spanish lexical decision task, strengthening the assertion that phonological coding of the primes takes place relatively early in the word recognition process.

Interestingly, the importance of phonological codes in word identification has been demonstrated in both orthographically shallow languages, where there is a direct mapping between orthography and pronunciation (e.g., Serbo-Croatian, for a review, see Carello, Turvey, & Lukatela, 1992) and orthographically deep languages, where the mapping appears to be more arbitrary (e.g., Chinese, for a review, see Tan & Perfetti, 1998). Clearly, phonological information can constrain visual word recognition even in logographic scripts where one would expect meaning to be derived directly from ideograms (Hoosain, 1991). For example, Tan and Perfetti (1999) sequentially presented pairs of Chinese words in a meaning-judgment task, in which subjects were asked to judge whether the two words had the same meaning or not. On trials where participants were supposed to make a “no” judgment (i.e., the two words had different meanings), the “no” response had longer latencies when the foil was homophonous with the base word compared to when it was not.

There have been additional tasks used to investigate the early influence of phonological processes. For example, Van Orden (1987; Van Orden, Johnston, & Hale, 1988) used a semantic categorization task, in which subjects had to decide whether a given word was a member of a semantic category. The intriguing finding here is that subjects produced considerably higher error rates for words that were homophones of an exemplar (e.g., MEET for the category FOOD), compared to an orthographically related control (e.g., MELT). This finding suggests a clear role of phonological information in accessing the semantics necessary for category verifications, and nicely converges with the results from the Tan and Perfetti (1999) study with Chinese characters. Jared and Seidenberg (1991) replicated this pattern showing that this effect is more likely to occur for low-frequency words. This pattern also appears to be consistent with the earlier observation of an interaction between frequency and spelling-to-sound regularity that was observed in word pronunciation performance (also, see Rodd, 2004). In another paradigen, Ziegler, Ferrand, Jacobs, Rey, and Grainger (2000) used an incremental priming technique, by manipulating the duration of the prime, which provides a window into the time-course of masked priming effects. They found clear orthographic and phonological priming effects in both naming and lexical decision performance, with the naming task being more dependent upon phonological priming. This study is particularly noteworthy because it provides a method to help understand the temporal locus of such priming effects. Finally, it is also worth noting that just as in the case of orthographic priming, there is also evidence of phonological priming in the parafoveal priming paradigm in more natural reading contexts. Specifically, Pollatsek, Lesch, Morris, and Rayner (1992) found that previews that were homophones of targets (e.g., site-cite) facilitated performance (both in pronunciation latencies and fixation durations during reading), compared to nonhomophonic previews that were controlled for orthographic similarity (e.g., cake-sake). Lee, Binder, Kim, Pollatsek, and Rayner (1999) have extended this work with a fast-priming paradigm (for a description, see Sereno & Rayner, 1992), a task which taps early stages of word processing. They observed an interesting prime by word frequency interaction; specifically, homophonic priming was primarily obtained with high-frequency word primes. Taken together, these findings not only support the role of phonology as an access code, but also suggest that lexical information may be guiding phonological coding early in fixations during reading (Lee et al., 1999).

3) “Semantic” Priming Effects

The semantic (associative) priming paradigm is clearly the most studied area of priming. (Because of space limitations, the present section will be limited to single word priming studies, see Morris, this volume, for a review of sentential semantic priming effects.) This enterprise began with a seminal study by Meyer and Schvaneveldt (1971). They found that subjects were faster to make lexical decisions to word pairs when the words were related (e.g., CAT-DOG) compared to when the words were unrelated (e.g., CAT-PEN). The prevailing zeitgeist was ready to welcome such a finding for a number of reasons: first, the dependent measure was response latency and response latency measures were becoming the mainstay of cognitive experiments. Second, the study nicely demonstrated top–down contextual influences (e.g., semantic relations) on what appeared to be a bottom up, stimulus driven word recognition processes. This was a major emphasis in Neisser's (1967) Cognitive Psychology that was published a few years earlier. Third, the effect was quite robust and easily replicated. Fourth, the semantic priming task appeared to be ideally suited to map out the architecture of meaning-level representations and the retrieval operations that act upon such representations; both of these issues would at least appear to be critical to higher-level linguistic performance.

a) Semantic or associative effects?

There is little controversy that across the major tasks used to build word recognition models (threshold identification, lexical decision, pronunciation, and on-line measures of eye-movements during reading), words are better recognized when embedded in semantically related contexts compared to unrelated contexts. However, there are many questions that have arisen regarding this effect. For example, one might ask if the effect is truly "semantic" (i.e., reflects similarity in semantic features, Smith, Shoben, & Rips, 1974 or category membership, Collins & Quillian, 1969), or if it primarily reflects associative relationships among items. For example, DOG and CAT share a semantic and associative co-occurrence relationship, whereas RAT and CHEESE appear to primarily share an associative relationship. Two recent reviews of this topic appear to come to somewhat different conclusions. Lucas (2000) argued that there was indeed evidence that semantic priming effects truly reflected "semantic" information, whereas, Hutchison (2003) concluded that, with a few exceptions, a simple associative account could handle most of this literature. Of course, teasing apart semantic influences from associative influences has been rather difficult because these relationships typically co-occur. In an attempt to address this issue, researchers have attempted to identify items that are of the same category (e.g., glove-hat) but do not entail a strong associative relation, e.g., are not produced in associative production norm studies in which subjects are asked to generate associates to a given word (see, for example, Palermo & Jenkins, 1964). The results from three such studies (e.g., Lupker, 1984; Schreuder, Flores d'Arcais, & Glazenborg, 1984; Seidenberg, Waters, Sanders, & Langer, 1984b) indicate that there is still some priming with such stimuli in both lexical decision and in pronunciation, although the pure semantic effects are somewhat smaller in pronunciation.

One must be cautious in accepting the conclusion that there are pure nonassociative semantic priming effects. This caution is warranted for the following reasons: first, and foremost, it is unclear whether the relatively small, but "pure," semantic priming effects might be due to some lingering associative-level relationship for words that researchers believe only have a semantic relationship (e.g., GLOVE-HAT are probably more likely to co-occur compared to the pair GLOVE-PEN). Second, as noted below, there is evidence that priming can occur across mediated pairs within the memory network. Thus, it is at least possible that some of the priming from GLOVE to HAT is due to GLOVE priming CLOTHES and CLOTHES priming HAT. Third, when one considers lowcategory dominance pairs, words that are categorically related but may have little associative relationship, one finds that there is relatively little priming in pronunciation performance (Keefe & Neely, 1990; Lorch, Balota, & Stamm, 1986); however, in lexical decision performance, there appears to be equivalent priming for high- and low-category dominance pairs (e.g., Lorch et al., 1986; Neely, Keefe, & Ross, 1989). The difference between pronunciation and lexical decision performance is particularly noteworthy here. As noted below, a number of researchers have suggested that at least part of the priming effect observed in the lexical decision task may be due to a type of post-lexical checking process. Subjects can use the relatedness between the prime and target to bias their "word" response because nonwords by definition are never semantically related to the primes. In fact, Neely et al. (1989) have found that the priming effect for low-dominance exemplars in the lexical decision task depends upon the ratio of nonwords to words. Neely et al. argued that the nonword/word ratio should modulate the likelihood of the checking process being engaged in the lexical decision task. Hence, because of the taskspecific list context effect in this study (i.e., the effect of the nonword/word ratio), one may question the argument for a pure semantic priming effect in access processes (also see Balota & Paul, 1996). In the following discussion, we will use the term "semantic" priming effects, however, the reader by now should understand that many of these effects could be primarily "associative" in nature.

b) Mediated priming effects

At an intuitive level, the finding that subjects are better at recognizing words that are embedded in related contexts compared to unrelated contexts is no great surprise. (Although, as described below, it is not so intuitive what mechanisms are responsible for such effects.) However, the priming literature has also provided some very counterintuitive findings. Consider the two words LION and STRIPES. These two words do not have any obvious direct relation, but do have an indirect relation through the word TIGER. Such items have been referred to as mediated pairs and the research addressing mediated priming effects has provided some interesting results. First, in a standard lexical decision task in which subjects only respond to the target string, there is little evidence for mediated priming (cf. Balota & Lorch, 1986; de Groot, 1983; den Heyer, Sullivan, & McPherson, 1987). However, if one changes the lexical decision task so that subjects either (a) make lexical decisions about the prime and target (McNamara & Altarriba, 1988) or (b) only make a response to word targets and not respond to nonword targets (den Heyer, Sullivan, & McPherson, 1987), mediated priming does occur in the lexical decision task. Moreover, when one now turns to the pronunciation task, one does find mediated priming effects (Balota & Lorch, 1986). Researchers have again argued that checking processes tied to the lexical decision task can strongly control when mediated priming effects will be found in this task (e.g., Balota & Lorch, 1986; McNamara & Altarriba, 1988; Neely, 1991). The notion is that checking for a relationship between the prime and target will not yield a successful outcome for mediated prime–target pairs, because such pairs do not share any obvious relationship. Thus, a negative outcome from the checking process may override the mediated influence from the prime to the target.

c) Threshold priming effects

A second important finding in this literature deals with threshold semantic priming effects, mentioned earlier. In the initial studies in this area, researchers first determined each subject's threshold wherein he or she can no longer discriminate between the presence or absence of a stimulus. These thresholds are then used in a later semantic priming task, in which the prime is presented at a subject's threshold and the target is presented in a lexical decision task. The intriguing finding here is that there still is evidence for semantic priming effects, under conditions in which subjects apparently can no longer make presence/absence decisions about the prime item (Balota, 1983; Carr & Dagenbach, 1990; Dagenbach, Carr, & Wilhelmsen, 1989;

Fowler, Wolford, Slade, & Tassinary, 1981; Marcel, 1983; Marcel & Patterson, 1978). There have also been similar findings reported in the pronunciation task (Carr, McCauley, Sperber, & Parmelee, 1982; Hines, Czerwinski, Sawyer, & Dwyer, 1986). Although there is some concern regarding whether subjects are truly at an objective presence/absence threshold (see Cheesman & Merikle, 1984; Holender, 1986; Merikle, 1982), it is clear that primes presented under very degraded conditions still produce semantic priming effects. It is noteworthy that the threshold priming literature has also been extended to functional neuroimaging techniques. For example, in an event-related potential/functional magnetic resonance neuroimaging study, Dehaene et al. (1998) used number primes that were so briefly presented that participants were unable to discriminate them from foils. Nevertheless, these primes influenced performance on a semantic comparison task (press one key if the target is less than 5 and another key if the target is greater than 5), and modulated hemodynamic measures of brain activity. As in the mediated priming studies, these studies indicate that conscious access to a prime–target relationship does not appear to be a necessary condition for obtaining semantic priming effects.

In some studies of threshold priming, stimuli and/or targets are repeated across trials, with thresholds being carefully monitored. There has been a recent debate about whether such effects in these paradigms reflect unconscious access to meaning at the whole-word level (Abrams & Greenwald, 2000; Damian, 2001; Naccache & Dehaene, 2001). For example, Abrams and Greenwald (2000) argued that threshold priming effects in these studies may reflect automatized stimulus–response mappings that develop as participants make responses to visible targets across trials (for an alternative view, Damian, 2001, but also see Kunde, Kiesel, & Hoffmann, 2003). Specifically, after participants repeatedly (and consciously) classify smut and bile as negative words, smile (smut-bile hybrid) subsequently functions as a negative valence masked prime (Abrams & Greenwald, 2000). No significant priming is found for masked primes that had not earlier appeared as a target to be classified. These findings question the traditional premise that threshold primes are analyzed at the whole-word level, and suggest that subconscious processing may instead involve sublexical analyses. There is, however, some recent evidence that these findings are specific to words. With numbers, masked primes are apparently able to provide access to long-term semantic memory (Greenwald, Abrams, Naccache, & Dehaene, 2003). The important point here is that one needs to be cautious in interpreting "threshold" priming effects when stimuli are repeated across trials.

It is also noteworthy that the masked priming paradigms have been extended to the domain of social psychology. The overarching question of interest is whether affective states can be automatically triggered by threshold-level primes. For example, Fazio, Sanbonmatsu, Powell, and Kardes (1986) found evidence for automatic attitude activation using an adjective connotation task (i.e., rate a target as "good" or "bad"). They observed that participants rated a negative valenced target (e.g., DISGUSTING) more quickly when it was preceded by a negative prime (e.g., COCKROACH), compared to a control item. Wittenbrink, Judd, and Park (1977) found similar priming effects under highly masked prime conditions. In a highly cited paper by Devine (1989) using the Neely (1977) automatic and controlled distinction in semantic priming, there was clear evidence of automatic activation of racial prejudice at short SOAs that was ultimately controlled at longer SOAs (see also Payne, 2001; Lambert et al., 2003). In reviewing this literature, Fazio (2001) has argued that such attitude priming is automatic and unconscious (also see De Houwer, Hermans, & Eelen, 1998).

Automatic influences from masked primes have also been detected using more ecologically valid paradigms. Bargh and Chartrand (1999) provide a comprehensive review of this literature. For example, Bargh, Chen, and Burrows (1996) found that participants presented with highly masked primes that presumably activated "rudeness" traits (e.g., rude, impolite, obnoxious) were more likely to interrupt a subsequent conversation than if they were primed with "politeness" traits (e.g., respect, considerate, polite). Collectively, the evidence from attitude and affect priming in social psychology is in-line with the evidence from semantic masked priming in visual word recognition. Given the cascadic nature of the models that we discussed earlier, such a pattern might be expected. However, this literature also clearly demonstrates that one needs to be cautious and use converging evidence to evaluate whether such effects are in the purest sense unconscious.

d) Backward priming effects

The third area that is counterintuitive is backward priming. There are two types of backward priming effects. First, there is evidence (Balota, Boland, & Shields, 1989; Kiger & Glass, 1983) that indicates one can still find semantic priming (DOG-CAT vs. PEN-CAT) even when the prime (DOG or PEN) is presented temporally after the target (CAT). These results suggest that early on in target processing, subsequent related prime information/activation can actually “catch-up” to influence response latencies to the target. Such an effect would appear to most naturally fall from a cascadic framework in which partial activation is released from representations before such representations have reached threshold.

A second type of backward priming effect is backward semantic priming. In backward semantic priming, prime–target pairs are presented that entail directional relations, e.g., BELL is related to BOY in the BELL-BOY direction, but not in the BOY-BELL direction. Koriat (1981) and Seidenberg et al. (1984b) have reported evidence of backward priming in the lexical decision task. However, when one turns to the pronunciation task, there is relatively little evidence of backward priming (Seidenberg et al., 1984b), except under short stimulus onset asynchronies (see Kahan, Neely, & Forsythe, 1999; Peterson & Simpson, 1989). It is possible that at short SOAs, there is sufficient temporal overlap between the target and the context to produce the first type of backward priming, noted above, even in naming.

4) Syntactic Priming

If associative/semantic context does indeed influence lexical processing, then it is quite possible that syntactically appropriate vs. inappropriate contexts might also influence lexical processing. In fact, effects of syntactic context on word recognition might be quite informative. At one level, one might argue that associative pathways between syntactically appropriate words might be represented within the lexicon, simply due to associative co-occurrence of such pairs (c.f., Ratcliff & McKoon, 1988). Likewise, one might argue that syntactic tags within lexical representations might produce priming to consistent syntactic representations. Alternatively, one might argue that syntactic representations are only engaged after word recognition and hence one might not expect syntactic priming effects in word recognition tasks.

One of the first syntactic priming studies was reported by Goodman, McClelland, and Gibbs (1981). Goodman et al. found that subjects were faster to make lexical decisions to targets (e.g., oven) that followed syntactically appropriate primes (e.g., my) compared to syntactically inappropriate primes (e.g., he). Seidenberg et al. (1984b) replicated this pattern in a lexical decision task, but only obtained marginal effects in the pronunciation task. As in the priming studies mentioned above, Seidenberg et al. argued that the syntactic priming effect in the lexical decision task was probably due to some post-lexical processing of the relation between the prime and target. At first, it appeared that Seidenberg et al.'s arguments are not totally correct, because West and Stanovich (1986) obtained relatively large syntactic priming effects in both the pronunciation task and the lexical decision task. However, Sereno (1991) argued that the past syntactic priming studies have used relatively long prime–target SOAs, and hence may be due to attentional expectancies. In a series of studies, with highly masked primes, Sereno found clear syntactic priming effects in lexical decision that were eliminated in naming, consistent with the Seidenberg et al.'s original arguments about taskspecific post-lexical checking processes producing the syntactic priming effects.

5) Prime Type by Factor Interactions

Of course, the importance of the semantic priming literature is not simply the demonstration that certain factors produce facilitation in the lexical decision and naming tasks, but its importance extends to the intriguing interactions that have been uncovered. As an example, consider the following intriguing pattern of interactive effects: (a) semantic priming effects are larger for low-frequency words than for high-frequency words (Becker, 1979); (b) semantic priming effects are larger for degraded words compared to non-degraded words (Becker & Killion, 1977; Borowsky & Besner, 1991); (c) there are additive effects of stimulus degradation and word frequency (see Balota & Abrams, 1995; Becker & Killion, 1977; Borowsky & Besner, 1991). Traditionally, this constellation of findings has been used to support independent, sequentially organized stages in lexical processing (Borowsky & Besner, 1993; Plourde & Besner, 1997; Sternberg, 1969). In contrast, Plaut and Booth (2000) have argued that a single-mechanism PDP model, implemented with a sigmoid activation function, can more parsimoniously simulate these effects, along with additional findings in the literature. This debate has recently resurfaced, with Borowsky and Besner (2005) contending that there is insufficient evidence that the PDP model implemented by Plaut and Booth (2000) can simultaneously achieve high lexical decision accuracy and correctly simulate the joint effects of stimulus quality, word frequency, and priming in speeded lexical decision. Instead, they argue that the available evidence is more consistent with serially organized processing stages that are differentially sensitive to degradation, semantic relatedness, and word frequency. Evidence for independent stages of processing is especially intriguing when considering the human word recognition architecture.

6) Theoretical Accounts of Semantic Priming Effects

The importance of the semantic priming paradigm has not simply been restricted to models of word recognition, but also has extended to more general issues concerning representation and retrieval processes. We shall now briefly discuss some of the theoretical issues that have been nurtured by this literature, but the interested reader should see Neely (1991), Hutchison (2003), and McNamara (2005) for a full discussion of these theoretical mechanisms.

a) Automatic spreading activation

The notion that semantic/lexical memory may be represented by nodes that reflect concepts and that such conceptual nodes are interconnected via associative/semantic pathways has been central to a number of developments in cognitive psychology (e.g., Anderson, 1976, 1983; Collins & Loftus, 1975; Posner & Snyder, 1975). As Anderson (1983) points out, the spreading activation metaphor has probably been most strongly supported by the semantic priming paradigm. When a node in memory becomes activated via stimulus presentation or via internal direction of attention, the notion is that activation spreads from that node along associative pathways to nearby nodes. Thus, the reason that subjects are faster to recognize DOG when it follows CAT, compared to when it follows PEN is because the underlying representation for these two words are connected via an associative/semantic pathway and when CAT is presented activation spreads from its underlying node to the node underlying DOG. Thus, the representation for DOG needs less stimulus information to surpass threshold.

Although there is a limited capacity version of spreading activation theory (e.g., Anderson & Bower, 1973), by far, most of the work in the priming literature has addressed the automatic nature of the spreading activation mechanism. In one of the clearest expositions of this mechanism, Posner and Snyder (1975) argued that the automatic spreading activation mechanism was (a) fast-acting, (b) independent of subjects' conscious control, and (c) primarily produces facilitation for related targets and little inhibition for unrelated targets, compared to an appropriate neutral baseline condition (see Neely, 1977). Because of controversies regarding the adequacy of a given neutral prime condition (see, for example, Balota & Duchek, 1989; de Groot, Thomassen, & Hudson, 1982; Jonides & Mack, 1984; Neely, 1991), we will primarily focus on Posner and Snyder's first two characteristics.

There are a number of important semantic priming results that would appear to support Posner and Snyder's automatic spreading activation mechanism. First, the evidence for semantic priming under highly masked priming conditions, reviewed above, is consistent with the notion that priming effects are independent of consciously controlled processing (e.g., Balota, 1983; Dehaene et al., 1998; Fowler et al., 1981; Marcel, 1983). Second, the evidence that there are mediated priming effects at relatively short prime–-target SOAs (e.g., from LION to STRIPES), when it is unlikely that subjects have sufficient time to generate an attentional expectancy for the mediated target also supports the notion of an automatic spread of activation within a memory network. Finally, the findings that prime-expectancy instructions (Neely, 1977) and relatedness proportion manipulations have relatively little impact at short SOAs (den Heyer, Briand, & Dannenbring, 1983), but strong influences at long SOAs, support the notion that the automatic spreading activation mechanism is relatively fast acting (i.e., occurs at short SOAs), decays quickly, and is independent of subjects' conscious expectations.

Although there appears to be support for something akin to an automatic spreading activation mechanism, there are some caveats. For example, initially, there was little evidence of priming effects occurring across unrelated words (e.g., facilitation from LION to TIGER in LION-CHALK-TIGER compared to FROG-CHALK-TIGER, e.g., Gough, Alford, & Holley-Wilcox, 1981; Masson, 1991; Ratcliff & McKoon, 1988). Clearly, if the effect is automatic, one would expect such effects. In this light, it is noteworthy that more recent studies by Joordens and Besner (1992), McNamara (1992), and Balota and Paul (1996) have obtained such priming effects. Of course, one might expect such priming effects to be relatively small because the unrelated word may have the effect of shifting attention away from the related prime and this shift may override any pure spreading activation effect. A second potential problem with the automatic nature of spreading activation is that semantic priming effects can be eliminated when subjects process the primes in a very shallow fashion, e.g., responding to whether a given letter is in the prime or an asterisk is beside the prime (e.g., Henik, Friedrich, & Kellogg, 1983; Smith, 1979; Smith, Theodor, & Franklin, 1983). Unless the shallow processing task eliminates processing of the prime at the lexical level, one should expect automatic spreading activation and semantic priming effects under shallow processing conditions (for further discussion of this issue, see Besner, Smith, & MacLeod, 1990). Finally, Balota, Black, and Cheney (1992) have shown that prime-expectancy instructions (e.g., subjects are instructed to expect exemplars from the TREE category when presented the prime METALS) can influence naming performance even at very short prime–target SOAs. Thus, although there is support of an automatic spreading activation mechanism involved in semantic priming tasks, it appears that we still do not fully understand the constraints under which this mechanism operates (for a recent discussion of the automatic nature of spreading activation, see Neely & Kahan, 2001).

b) Attentional/expectancy effects

A second mechanism that presumably underlies semantic priming effects is a more attention-based expectancy factor (Balota, 1983; Becker, 1980; Favreau & Segalowitz, 1983; Neely, 1976, 1977). Here, when the prime is presented subjects generate expectancies about potential candidate targets. When the expectancy is correct, facilitation occurs, however, when the expectancy is incorrect, inhibition occurs. This expectancybased model of priming falls naturally from the work of Posner and Snyder (1975) and Neely (1977), wherein, instructional manipulations and list context effects have larger influences at long SOAs (when expectancies have had time to be generated) than at short SOAs. Of course, at one level, the impact of an attentional-based expectancy mechanism should not be surprising because it simply reflects the probability of correctly predicting the target word when given the prime. The more intriguing work here is the specification of the parameters that modulate the expectancy effects, i.e., the rate at which expectancies are generated across time, the duration at which the expectancy is maintained, and the characteristics of such an expectancy set (for a detailed discussion of a semantic expectancy model, see Becker, 1980, 1985).

c) Backward-checking accounts

As noted above, a number of researchers have argued that priming effects in the lexical decision task may reflect influences at a post-lexical decision level (e.g., Balota & Lorch, 1986; de Groot, 1984; Forster, 1979, 1981; Neely, 1976, 1977; Neely & Keefe, 1989; Seidenberg et al., 1984b; Stanovich & West, 1983). Subjects can rely on finding a relationship between the prime and target to bias the "word" response in the lexical decision task, because nonwords are never related to the primes. This would have the effect of facilitating "word" decisions to related prime–target trials and possibly inhibiting "word" decisions to unrelated prime–target trials. As described above, there is considerable support for such a mechanism in the lexical decision task. For example, the finding that there is backward priming in the lexical decision task (e.g., priming from BOY to BELL) suggests that subjects can use the target to check in a backwards direction (BELL to BOY) for a potential relationship to the prime item. Although the backward checking mechanism would appear to be primarily a nuisance variable tied to the lexical decision task, one might argue that this checking process may reflect a tendency in natural language processing to integrate the meaning of the current word with the ongoing comprehension of the previous words (for a full discussion of the backward checking mechanism, see Neely & Keefe, 1989). As noted above, in support of this possibility, Kahan et al. (1999) found some evidence of backward checking at short SOAs even in naming performance.

d) Compound-cue model

Ratcliff and McKoon (1988) developed a model that takes a quite different approach to priming effects in the lexical decision task. The model is based on a formal model of episodic recognition memory developed by Gillund and Shiffrin (1984). In Ratcliff and McKoon's model, items in short-term memory serve as a compound cue with the more recently presented items having a larger influence on the output of the retrieval process. If the prime and target are associated then this will provide a higher familiarity value than if the prime and target are not associated. Familiarity is then used to predict response latency via a random-walk decision process (Ratcliff, 1978), wherein, high-familiar compound cues produce relatively fast "yes" decisions and low-familiar compound cues produce relatively slow "no" decisions. Intermediate values of familiarity produce relatively slower and less accurate decisions. Hence, if familiarity is modulated by the degree to which primes and targets are either directly associated or share associates in memory, then one should find that related prime–target pairs will produce higher familiarity values and faster response latencies than unrelated prime–target pairs.

Although the compound cue model does provide an interesting alternative to prime-induced mechanisms, there are some limitations to this approach. For example, the model is primarily a model of the lexical decision task, and hence, does not account for the wealth of interesting priming data from the pronunciation task, along with other tasks. Neely's (1991) tripartite (spreading activation, attentional expectancies, and backward checking) framework accounts for both lexical decision and pronunciation results by assuming logogen-type word recognition devices that are also connected to a phonological output system used for pronunciation. Second, and more importantly, the distinction between the compound cue model and the spreading activation framework may be more apparent than real. In both frameworks, it is necessary to map the influence of relationships between words onto priming effects. Within the spreading activation framework, this mapping involves the preactivation of related concepts in memory, whereas, within the compound cue model, this mapping is based on a rule that computes familiarity based on associations within long-term memory. At this level, the major distinction between the spreading activation framework and the compound cue model involves this mapping process.

e) Plaut and Booth's (2000) Single-mechanism connectionist model

In contrast to Neely's tripartite framework described above, Plaut and Booth have claimed that a distributed network model can account for semantic priming lexical decision phenomena using a single mechanism. Implementing a distributed attractor network with distributed orthographic and semantic representations (Plaut, 1995), Plaut and Booth were able to account for a number of theoretically interesting findings, including the surprising observation that only participants with high perceptual ability exhibited the priming by frequency interaction (i.e., greater priming for low-frequency words); participants with low perceptual ability showed equal priming for both high- and low-frequency targets. Like the Seidenberg and McClelland (1989) model, however, the connectionist view of priming faces challenges. For example, as mentioned earlier, there is an ongoing debate about whether semantic priming is better accommodated by a single-mechanism account or by separate mechanisms that invoke distinct sets of computational principles (see Borowsky & Besner, 2006). Nevertheless, this work represents an interesting advance in that it includes a computationally implemented architecture that has been applied across a number of cognitive domains and accommodates some intriguing data in the priming literature and takes a step toward tackling the important topic of individual differences.

f) Masson's (1995) distributed memory model of semantic priming

Masson's model, based also on distributed connectionist principles, provides a framework for accommodating semantic priming in speeded naming that neither appeals to spreading activation nor compound cues. In this model, conceptual knowledge is represented via distributed orthographic, phonological, and semantic units that are connected by weighted pathways. Importantly, Masson's network, a Hopfield (Hopfield, 1982) net variant, does not distinguish between input, hidden, and output units. The basic principle in the model is that semantically related words have very similar patterns of activation in the semantic units. When a semantically related prime is presented, activation in the semantic units starts moving toward a pattern that is similar to the pattern of activation of the to-be-presented target. When the target appears, the overlap between its pattern and the pattern of activation in the semantic units helps the phonological units converge more rapidly on the target's pattern, and hence, speeds naming responses. This model is able to account for the intervening stimulus effect, which, as mentioned above, is the observation that interpolating an unrelated word between the prime and the target reduces the priming effect in naming performance, a finding that the spreading activation framework does not readily predict. However, it is also the case that this model has not been extended to the wealth of data that Neely's tripartite framework appears to be able to handle.

7) Summary of Context/Priming Effects

The priming literature has provided an extremely rich data base to develop models of context effects, memory retrieval, and word recognition. Because of space limitations, we were unable to provide a review of other important models of semantic priming effects such as Becker's (1980) verification model, Norris' (1986) plausibility-checking model, and Forster's (1976) bin model. Each of these models provides intriguing alternative perspectives on semantic priming effects. At this point in theory development, it appears that no single model of priming readily accounts for the richness and diversity of this literature, and it would appear that multiple mechanisms will need to be postulated to account for the breadth of semantic priming effects.

7. Attentional Control, Modularity, and Time Criterion Models

The models reviewed earlier appear to have a relatively passive architecture wherein different systems accumulate information across time. However, in some instances, it may be advantageous for the reader to modulate the contribution of a given pathway depending upon the task demands or particular reading context. For example, one might expect different emphases on distinct systems when proofreading, comprehending, or checking for grammaticality. Virtually, every theory of word recognition posits multiple ways of accessing or computing the phonological code from print. In the DRC model, one can compute a phonological code via the lexical route, which maps the whole word onto a lexical representation to access phonology, or via the sublexical route, which computes the phonology via the spelling-to-sound correspondences in the language; in PDP models, the phonology can be computed by differential emphasis on the direct orthographic to phonological connections or the indirect connections via semantics. The question naturally arises whether there is any control of which processing pathway influences performance in a given task. This is important because it brings into question the modularity of the lexical processing system (see Fodor, 1983).

One way to examine the control issue is to present words that place different demands on the lexical and sublexical information. For example, within a DRC model, nonwords should bias the sublexical pathway. However, low-frequency exception words should bias the lexical pathway, since the sublexical pathway would lead to regularization errors for low-frequency exception words, i.e., pronouncing pint such that it rhymes with hint. Monsell et al. (1992) found that naming latencies to high-frequency irregular words were faster and more accurate when embedded in lists with other irregular words, than when mixed with nonwords. Monsell et al. suggested that exception word context directed attention to the lexical pathway, which is more appropriate for naming exception words, than the sublexical pathway. Additional studies have found similar influences of pathway priming (e.g., Rastle & Coltheart, 1999; Reynolds & Besner, 2005b; Simpson & Kang, 1994; Zevin & Balota, 2000).

Although intuitively appealing, the evidence for route priming has been quite controversial. Specifically, work by Kinoshita and Lupker (2002, 2003a) suggests that much of the earlier findings can be accounted for by a time criterion model. The time criterion perspective is important in a number of domains so we will briefly review it here. Specifically, there is evidence that participants adopt a time criterion whereby they are likely to produce a response at a latency that is biased toward the average of the latencies in a block of trials. Consider the word-frequency effect (presumably a reflection of the lexical route). In two pure independent blocks, assume that a set of low-frequency words produces response latencies on the average of 700ms and a set of high-frequency words produces response latencies on the average of 600ms. If one now embeds these same words in the context of nonwords that produce an average response latency of 700ms, the word-frequency effect will likely diminish. That is, latencies to the low-frequency words will remain relatively the same (because the latencies for both low-frequency words and nonwords are 700ms), whereas latencies to the high-frequency words will increase considerably, i.e., migrate toward the time criterion invoked by mean latency of the nonwords. Hence, the word-frequency effect will decrease in the context of nonwords not because of a decreased reliance on the lexical pathway, but rather because of a change in the temporal criterion to produce a response.

The evidence clearly suggests that participants do adopt a time criterion based on the difficulty of items within a block. However, there is also evidence that appears to be consistent with a pathway control perspective above and beyond the time criterion effects. For example, all of the effects reported by

Zevin and Balota (2000) hold even after the response latencies to the context items are partialed out via analyses of co-variance. Of course, if the time criterion model were the only responsible variable in this study, one should not find this pattern. Moreover, Kinoshita, Lupker, and Rastle (2004) have recently provided evidence that one can indeed modulate the lexicality effect (words faster than nonwords) via list context manipulations. However, they were unable to modulate the regularity effect (regular words faster and/or more accurate than exception words) by list context manipulations. In addition, Reynolds and Besner (2005b) have recently demonstrated that one can find lexical and sublexical pathway switching above and beyond any response latency criterion effects. Although there is accumulating evidence for some level of pathway control, further work is clearly necessary in this area. Indeed, the extent to which attentional systems modulate the information in distinct pathways has important implications for future modeling endeavors, and quite naturally would accommodate taskspecific influences that have been emphasized in the present chapter. Moreover, time criterion perspectives are important in understanding how the word recognition system adjusts to the local constraints of an experiment and may have important implications for other cognitive paradigms that rely on response latency measures. At this level, time criterion effects may be viewed as an example of attentional control.

8. Developments of New Approaches and Analytic Tools to Guide the Journey from Features to Meaning

In the following sections, we will describe some recent developments in approaches to studying word recognition. Again, this is not a comprehensive review, but simply a brief summary to expose the reader to some of the interesting techniques that are helping researchers constrain how humans process visual words.

1) Neuroimaging Techniques

In the past decade, tremendous advances in neuroimaging methodology have provided another window into the dynamics of lexical processing (also see Just and Mason, this volume). Specifically, neuroimaging techniques like positron emission tomography (PET), functional magnetic resonance imaging (fMRI), and event-related potentials (ERPs) now allow researchers to localize and measure the time course of activity of brain regions that are recruited by a particular cognitive task (Fiez & Petersen, 1998). One particularly exciting development has been the advent of event-related fMRI designs (Dale & Buckner, 1997). In PET studies and early fMRI studies, blocked designs (i.e., experimental conditions are blocked) were mandatory, making paradigms like the lexical decision task impractical. Event-related fMRI allows researchers to extract the fMRI BOLD (blood oxygen level dependent) response for specific trials and to conduct standard word recognition experiments in the scanner.

As Fiez and Petersen point out, neuroimaging allows one to make both coarse as well as refined fractionations of brain regions that are involved in reading. For example, Petersen, Fox, Posner, Mintun, and Raichle's (1988) seminal study elegantly demonstrated that remarkably distinct brain regions were activated by different levels of single-word processing. Specifically, at a relatively coarse level, PET scans revealed that compared to appropriate baseline conditions, occipital areas were active for passive viewing of words (orthography), temporal areas were active for reading words aloud (phonology), and frontal regions were active when participants generated verbs to nouns (semantics). By varying tasks demands and contrasting neural activation in reading aloud versus a control condition, these researchers were able to identify broadly the functions of different regions.

More recently, research designs have been employed to make finer-grained differentiations of regions that support different reading operations. For example, in a PET study of speeded naming, Fiez, Balota, Raichle, and Petersen (1999) manipulated the following three variables: lexicality (word vs. pronounceable nonword), frequency (high vs. low), and spelling-to-sound consistency (consistent vs. inconsistent). As discussed earlier, these variables have been central in the developments of models of word recognition, and so it is useful to explore the underlying circuitry. Fiez et al. (1999) found a number of noteworthy effects. First, a left frontal region showed effects of consistency and lexicality, indicating that this area may be involved in orthographic-to-phonological transformation. Second, there was greater activation for low-frequency words in a left temporal region and the supplementary motor area, which implicate these regions in the access and storage of lexical-level information. Third, effects of consistency were found bilaterally in the primary motor cortex, suggesting that consistency may influence both recognition and motor production systems; this surprising constraint has yet to be considered by extant theories of word recognition (but see, Kawamoto, Kello, Jones, & Bame, 1998). Fourth, the left inferior frontal gyrus showed a pattern analogous to the behavioral frequency by regularity interaction discussed earlier. Specifically, just as naming latencies are particularly slow for low-frequency inconsistent words compared to low-frequency consistent words, high-frequency inconsistent words, and high-frequency consistent words (Seidenberg et al., 1984a), the left inferior frontal gyrus showed strong activation only for low-frequency inconsistent words. This study demonstrates how manipulating stimulus properties in a neuroimaging paradigm can be used to complement and extend theoretical accounts that have hitherto been based on behavioral data.

In the remainder of this section, we will review some recent neuroimaging studies and discuss how these studies contribute to our understand of word recognition. Obviously, due to space constraints, this review is selective. Also, rather than enumerating in minute detail which brain regions are activated by which task, we will be using more sweeping brushstrokes to describe the functional neuroanatomy of reading.

a) Is there convergence across studies?

A reasonable concern one may have regarding neuroimaging research is the extent to which findings generalize across laboratories and studies. Variability across studies may arise as a result of intersubject variability and slight differences in methodology, making it difficult to establish consistent regions of activation (Turkeltaub, Eden, Jones, & Zeffiro, 2002). A few articles have attempted to review results from multiple studies in order to answer this question. For example, Fiez and Petersen (1998) reviewed nine studies where participants read aloud single words, and found encouraging convergence between studies. Basically, they combined the data across the studies by merging foci from different experiments into a single figure (Figure 9).

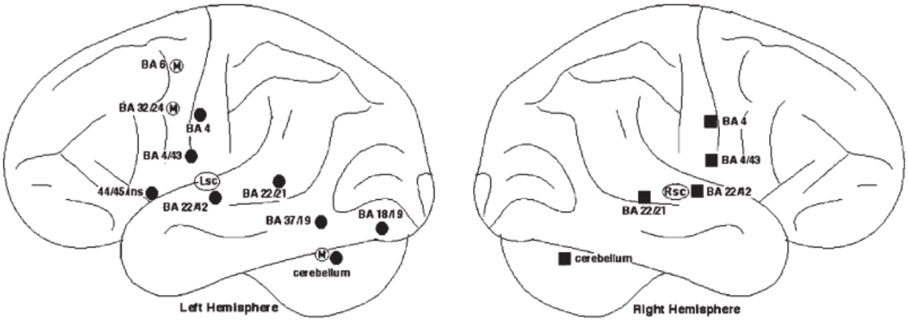


Figure 9. Brain regions that are consistently activated during word reading (Fiez & Petersen, 1998).

Fiez and Petersen (1998) established that a set of areas are consistently active during word reading, including the supplementary motor area, the cerebellum, the anterior cingulate gyrus (BA 32), left-lateralized fusiform and lingual gyri (BA 18 and BA 37), the left inferior frontal gyrus (BA 44/45), bilateral activation in the anterior and posterior regions of the superior temporal gyrus (near BA 22), and dorsal and ventral portions of the post-central gyrus (near BA 4). Interestingly, a more sophisticated meta-analysis of 11 PET studies generated similar findings (also, see Price, 2004). A statistical map of convergent foci by Turkeltaub et al. (2002) included the bilateral motor and superior temporal cortices, presupplementary motor area, left fusiform gyrus, and the cerebellum (see Figure 10). This map was successfully validated against new fMRI data of word reading, supporting the reliability of these findings.

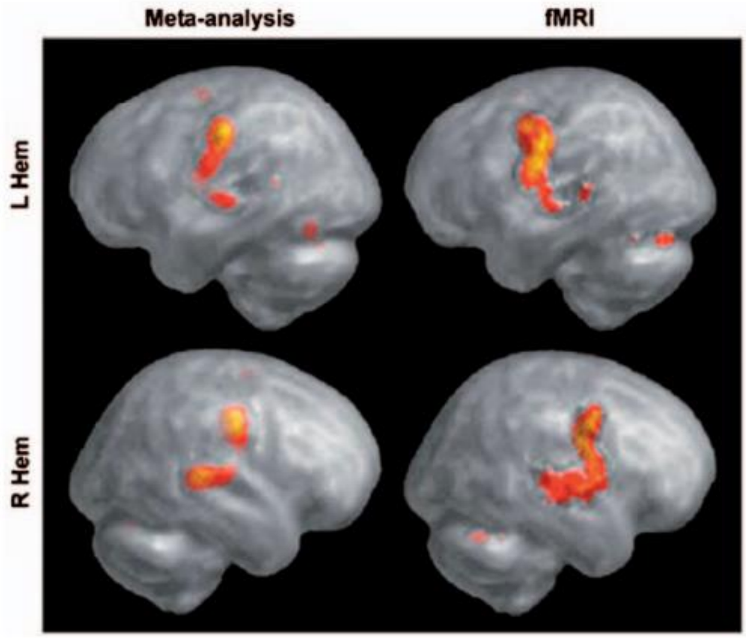


Figure 10. Statistical map generated by meta-analyses of 11 PET studies of word reading (Turkeltaub et al., 2002).

Taken collectively, the regions identified by neuroimaging are broadly compatible with the neuropsychological literature (Fiez & Petersen, 1998; but see Price, 2004, for exceptions). More importantly, these analyses generate candidate regions of interest that can be used by researchers to test new hypotheses.

b) Controversies regarding targeted areas

We have already discussed how the left inferior frontal gyrus is sensitive to spellingto-sound consistency manipulations (Fiez et al., 1999), implicating this region in processes that transform orthographic to phonological representations. This pattern has been nicely replicated in a number of other studies (see Herbster, Mintun, Nebes, & Becker, 1997; Rumsey et al., 1997). Note that some researchers have proposed another locus for the sublexical procedure, the left posterior superior temporal region (Simos et al., 2002), although this region may be associated more with phonological decomposition (Palmer, Brown, Petersen, & Schlaggar, 2004) than with orthographicto-phonological transformation per se.

Interestingly, there is far less agreement about the neural markers of lexical processing. Typically, frequency and lexicality effects have been used as indicants of lexical processing. Word-frequency effects may mark brain regions involved in the access and representation of either localist or distributed lexical-level information (Fiez et al., 1999). Although word frequency is easily the most studied variable in word recognition, it has received surprisingly little attention in the functional neuroimaging literature. The literature suggests greater activation for low-frequency words in a left temporal region (near BA 22) and the supplementary motor area (BA 6) in speeded naming (Fiez et al., 1999), and greater activation in the left inferior frontal gyrus (BA 44/45) and a number of subcortical structures in lexical decision1 (Fiebach, Friederici, Muller, & von Cramon, 2002).

A number of studies have also used lexicality effects (greater activation for words than nonwords) as a marker for the lexical processing (Binder et al., 2003; Fiebach et al., 2002; Ischebaeck et al., 2004; Rissman, Eliassen, & Blumstein, 2003; Simos et al., 2002). It seems plausible that greater activations for words reflect access to linguistic information, which may either be orthographic, phonological, or semantic (Fiez et al., 1999). Unfortunately, there seems to be little consensus on regions that show greater activation for words. Lexical decision studies have identified diverse regions, including the left angular gyrus (BA 39), left dorsal prefrontal cortex (BA 6/8), superior frontal gyri (BA 6/8/9), left rostral–ventral cingulate gyrus (BA 32/24), left posterior cingulate gyrus and precuneus (BA 23/29-31/7), and the junction of the left posterior middle temporal and inferior temporal gyri (BA 21/37). Curiously enough, in speeded naming, it is relatively difficult to find regions that show greater activations for words than nonwords. For example, in the Fiez et al. (1999) PET study, no region showed greater activation in the word condition compared to the nonword condition. This may be attributable to the lower spatial resolution of PET (compared to fMRI, but see Palmer, 2003), or to strategic effects induced by blocking. Using a more sensitive event-related paradigm with Japanese Kana words, Ischebeck et al. (2004) found greater activation for words in the left and right temporo–parietal areas (BA 39/40), the middle part of the left middle/inferior temporal gyrus (BA 21/20) and the posterior cingulate (BA 31).

The marked discrepancy between the lexical decision and naming findings may be partly attributable to differential demands of the two tasks; a theme that has consistently arisen in the current chapter. If indeed there are large task differences in the behavior, there clearly should be consequences for the neural underpinnings. For example, the meta-analyses we discussed earlier (e.g., Fiez & Petersen, 1998; Turkeltaub et al., 2002) were based on neuroimaging studies of naming. It will be interesting to see the degree to which statistical maps based on meta-analyses of lexical decision data show a similar pattern. This is a theoretically important question that has yet to be answered.

Nevertheless, in spite of the diversity of findings, two regions seem to be consistently associated with lexicality effects (word > nonword), the left angular gyrus (BA 39) and the left middle temporal gyrus (BA 21), a pattern which is nicely consistent with the Kana-naming study (Ischebeck et al., 2004). The left middle temporal gyrus has long been associated with language processing (Fiebach et al., 2002); other studies also implicate the region in the representation and processing of lexico–semantic (Pugh et al., 1996) and phonological word forms (Cohen et al., 2000). The left angular gyrus also seems to play a role in non modality-specific semantic processing (Binder et al., 2003).

PET and fMRI are not the only windows into the functional neuroanatomy of reading. While these measures have excellent spatial resolution, the intrinsic characteristics of these signals limit their temporal resolution. Event-related potentials (ERPs) have far more exquisite temporal resolutions, and so these measures are better suited to study the time course of word recognition processes. Scalp-measured ERPs reflect the brain electrical activity that is triggered by experimental stimuli, and capture in real-time cognitive processes on a millisecond basis (Kutas & Federmeier, 2000; Kutas and van Patten, this volume). Frequency effects, for example, are apparent in ERPs between 200 and 400 ms, and are most obviously correlated with a left anterior negative component called the lexical processing negativity (LPN) (King & Kutas, 1998). In a recent representative ERP study, Hauk and Pulvermüller (2004) investigated how word length and word frequency influenced the amplitude and peak latencies of event-related potentials in lexical decision. Early effects of length and frequency were observed, and the researchers interpreted the results as consistent with lexical access occurring as early as 150ms after the onset of visually presented words. More intriguingly, large length effects were observed in ERPs but not in the behavioral data; this dissociation demonstrates that psychophysiological measures may in some cases be more sensitive than behavioral data.

To recapitulate, in the foregoing discussion, we have briefly considered the neuroanatomical correlates of selected psycholinguistic effects. Clearly, this nascent work is exciting and informative, and many issues remain unexplored. In the final portion of this section, we will consider how neuroimaging has advanced our understanding of word recognition processes above and beyond traditional behavioral work.

c) What constraints are afforded by neuroimaging techniques?

It is incontrovertible that we know a great deal more about the neuroanatomy of language today than a mere 10 years ago. Nevertheless, it is also clear that neuroimaging of cognitive processes is still a relatively new area of investigation. Even though the conclusions we have presented are preliminary and may be revised not too far in the future, we would contend that neuroimaging data is an essential adjunct to response latency and accuracy data. Most obviously, neural correlates of behavior provide another level of explanation (Marr, 1982) that reveals how reading processes are physically instantiated in the brain. Moreover, Palmer et al. (2004) have cogently argued that collecting functional neuroimaging data affords two other important advantages. One, brain activation data can powerfully complement behavioral measures. For example, young and older adults may perform identically on a task (in terms of response latencies and error rates), but show marked differences in brain activity (also see Hauk & Pulvermüller, 2004; Schlaggar et al., 2002). In addition, neuroimaging data may also be useful in informing theories and adjudicating between competing models. For example, the DRC model (Coltheart et al., 2001) and the connectionist model (Plaut et al., 1996) adopt very different architectures for naming words that have been difficult to discriminate based on behavioral data, but it may be possible that converging evidence of the role of different neural substrates dedicated to specific operations may provide important information on how the brain implements the processes involved in word recognition.

2) Large-Scale Studies vs. Factorial Studies of Word Recognition

Word recognition researchers have traditionally employed factorial designs where item variables of interest (e.g., length, frequency, etc.) have been “manipulated,” and other factors known to affect performance have been controlled. This approach has been useful, but there are some limitations (see Balota et al., 2004). Recently, researchers have examined word recognition performance for large sets of words that are not constrained by selection factors, e.g., virtually all monosyllabic words (see Balota & Spieler, 1998; Besner & Bourassa, 1995; Kessler, Treiman, & Mullennix, 2002; Spieler & Balota, 1997; Treiman et al., 1995). Such datasets are useful in a number of ways. For example, using standard predictor variables, Balota et al. (2004) accounted for 49 and 50 percent of the variance in the lexical decision and speeded naming performance, respectively for a dataset of 2428 words. This is a multifold increase over current computational models (for a discussion of pros and cons for using accounted for variance as a critical variable in evaluating a model’s performance, see Balota & Spieler, 1998; Seidenberg & Plaut, 1998). This outcome was obtained despite the success these computational models have had in accounting for performance at the factor level. The large-scale item-level analyses provide another potentially important constraint in the evaluation of theoretical approaches to word processing. More recently, Balota and colleagues have collected naming and lexical decision latencies for over 40,000 words (Balota et al., 2002; Balota et al., in press). The English Lexicon Project website (http://elexicon.wustl.edu) provides a comprehensive data set of behavioral measures that researchers can easily access, via a search engine, along with a rich set of descriptive characteristics. Hopefully, this dataset will be helpful in extending current models to multisyllabic words, which as noted above is a potentially serious limitation in current models. Finally, as mentioned earlier, recent attempts to ground semantics in large scale natural databases of language use (e.g., Burgess & Livesay, 1998; Landauer & Dumais, 1997; Steyvers & Tenenbaum, 2005) have also been quite informative. Clearly, the computational power available today that affords analyses of these large-scale databases appears to be providing an important additional constraint on theory development.

3) RT Distributional Analyses

In standard word recognition experiments, one compares the mean response latency across several conditions to determine if the predictions generated by an experimental hypothesis are correct. Of course mean performance is not the only estimate of performance across a set of trials. Researchers have long noted that means of conditions are only one estimate available from performance. For example, in the standard Stroop task (i.e., naming the color that a word appears in), Heathcote, Popiel, and Mewhort (1991) provided a useful demonstration of how the shape of a response time distribution can provide useful information beyond estimates of central tendency. They found that the incongruent condition (e.g., the word blue appearing in the color red), compared to the neutral condition (e.g., the word block appearing in the color red), increased both the skewing and the central tendency of the reaction time distribution, but amazingly, the congruent condition (e.g., the word red appearing in the color red) increased skewing and decreased the central tendency, which basically masked any effect in means (for a replication of this pattern, see Spieler, Balota, & Faust, 1996). These researchers have fit reaction time distributions to ex-Gaussian functions, but other functions such as the Weibull or ex-Wald could also capture useful characteristics of the reaction time distributions. As theories become more precise regarding item level performance, there should be an increased level of sophistication regarding the predictions concerning the underlying reaction time distributions. For example, Balota and Spieler (1999) found that frequency and repetition influenced these parameters differently depending on the dependent measures, i.e., naming vs. lexical decision (however, see Andrews & Heathcote, 2001). Ratcliff et al. (2004) have recently used reaction time distributions to more powerfully test a diffusion model of lexical decision performance (see also Yap, Balota, Cortese, & Watson, in press). As models become more sophisticated, the precision of reaction time distribution analyses will be critical in their evaluation.

4) Individual Differences

Just as one may be losing information when averaging across items to estimate means, one is also losing information when averaging across individuals. Of course, there are standard comparisons of individual differences as a function of age, acquired or developmental dyslexia, or other neuropsychological impairment (see Perfetti, this volume), however, another possibility is that individuals may produce particular profiles of lexical processing. For example, if indeed the dual-route model is correct, one might find that some subjects rely more on lexical pathways, while other subjects rely more on sublexical pathways, and this could indeed be tied to the manner in which they were originally taught to read or inherent individual differences in capacities. The recent explosion of interest in differences in working memory capacity has been quite successful in identifying distinct cognitive processing profiles (see, for example, Engle, Kane, & Tuholski, 1999). With the advent of large datasets on individual subjects (see megastudies mentioned earlier) it is quite possible that such differences could be observed (for processing speed modulating the effects of orthographic neighborhood size, see Balota et al., 2004). Of course, this may also push researchers to more closely consider the reliability of effects, which at least within one domain, semantic priming, appear to be surprisingly low (see Stolz, Besner, & Carr, 2005).

9. Concluding Remarks

In the present chapter, we have attempted to provide the reader with an overview of the major issues addressed in the word recognition literature. To conclude, we would like to summarize some of the major themes that have spanned a number of the sections. First, in each of the sections, there has been evidence initially supporting a rather straightforward theoretical analysis and then there have been reports by "trouble-makers" that constrain the strength of the theoretical inferences available from a given task. For example, even in the word superiority paradigm, there have been arguments that partial information from the target letter could, in conjunction with the word-envelope, allow subjects to use a sophisticated guessing strategy to bias the correct choice (e.g., Krueger & Shapiro, 1979; Massaro, 1979). If this is the case, then the word-superiority effect may not reflect top–down impacts in perception, but rather, biases that occur at post-perceptual levels, based on partial information. Similar concerns were raised about the threshold identification, lexical decision, and pronunciation tasks. Of course, task analyses can be frustrating for theoreticians, however, before inferences can be made regarding the underlying locus or loci of a given variable, one should be especially careful in developing (or understanding) tasks that faithfully reflect such processes. Clearly, the adequacy of any theory rests on the quality of the tasks used to build that theory.

A second consistent theme that has surfaced in this review is whether there are separable analyses performed en route to word recognition or the apparent influences of multiple pathways are in large part merely a consequence of the activation and inhibition patterns across many lexical representations. Although some effects appear to be modeled quite well by interactive activation and parallel distributed processing systems, there have also been results that appear inconsistent with such systems. There are at least two likely outcomes to this area of work. First, more of the troublesome effects may fall from these models when networks that are closer to the size of an adult's vocabulary are implemented (see Seidenberg & McClelland, 1990). Second, it may be necessary to implement sublexical processing modules within such connectionist models to incorporate the strong evidence for multiple distinct access pathways. Clearly, this is still a central issue in the current state of model development (see Andrews, 2006).

A third theme in the present review is the type of statistical interaction that has been repeatedly observed. The vast majority of interactions in this literature are of the nature that Factor A has more of an effect at the level of Factor B that produces the slowest or least accurate performance. Consider for example word frequency. We have reviewed evidence indicating that compared to high-frequency words, low-frequency words produce larger effects of bigram frequency, spelling-to-sound consistency, word-body strength, concreteness, semantic priming, task (lexical decision task vs. category verification vs. pronunciation), repetition priming, neighborhood size, among others. There are at least two noteworthy aspects of these interactions. First, one may wish to argue that because of the development of automaticity, high-frequency words are recognized via routes that effectively bypass many sublexical stages of analyses. Hence, if one is interested in identifying many of the intriguing sublexical aspects of word recognition, one should primarily investigate the processing of low-frequency words. Alternatively, as Loftus (1978) has noted, on a simply statistical level, this particular type of interaction is one of the most difficult to interpret. In fact, it is possible, that if one considered percentage of overall response latency change as a function of the levels for Factor A and B, or a z-score transform of the data (taking into account variability), many of these interactions would disappear (for a detailed discussion of these issues, see Faust, Balota, Spieler, & Ferraro, 1999). Clearly, the assumption of a linear relations between response latency and underlying cognitive operations is a simplifying assumption, which will ultimately need to be faced by those studying the time-course of processes involved in visual word recognition, along with other cognitive operations.

In sum, we are hopeful that the reader agrees that at some level the word is to cognitive psychologists and psycholinguist as the cell is to biologists. Both entail many substructures and interact with many higher-level systems. The present overview of the word recognition literature may seem rather imposing, and sometimes it would appear that little progress is being made. However, this clearly is not the case; considerable progress has been made, especially within the last decade. Of course, the seductive simplicity of understanding lexical-level analyses surely is more apparent than real. As is often the case in a discipline, the more we know about a system, the more we develop procedures for generating and constraining our questions in the future. Given the new analytic methods that have come on line recently this will indeed be a very exciting next decade of research.

CONCLUSIONS

Given the literature reviewed here, showing a broad range of semantic effects in standard LDTs, I suggest that meaning-based infl uences are robust in visual word recognition. Similar conclusions were drawn by Balota et al. (2004), based on their examination of the effects of a number of semantic variables on visual word recognition performance. Balota et al. argued that their results were " … most consistent with a view in which meaning becomes activated very early on, in a cascadic manner, during lexical processing and contributes to the processes involved in reaching a suffi cient level of information to drive a lexical decision" (p. 312). The results reviewed here suggest that a number of different semantic dimensions are activated in lexical processing: these include the extent to which words evoke imagery, featural information, associated concepts, co-occurring concepts, bodily experience, multiple meanings, and affect. In the few cases where the effects of different dimensions of semantic richness have been directly compared (e.g., Mirman

& Magnuson, 2008 ; Pexman et al., 2008 ; Yap et al., 2011) the dimensions have each had independent relationships with word recognition behaviour. In fact, there is little evidence that these dimensions have.

Word recognition is the component of reading which involves the identification cation. In other words, it is how readers retrieve and select the right representation among others in the mental lexicon. Word recognition research has been central to work in cognitive psychology and psycholinguistics because words are relatively well-defined minimal units that carry many of interesting codes of analysis (i.e., orthography, semantics, syntax). The interest here is to define the perceptual unit in word recognition, it would seem obvious that the letter should be the primary unit of analysis in visual word recognition i.e. words are made by letters.

In this topic, we have described the major tasks employed, the different theoretical perspectives, many of the variables that influence word recognition performance, and some of the continuing controversies. Of course, this overview only provides a glimpse of the vast amount of research that has been conducted on visual word recognition. Although much has been accomplished, there is clearly need for continuing work in clarifying the processes engaged in the seductively simple act of visual word recognition

[This page is intentionally left Blank]

CHAPTER 8
LEXICAL PROCESSING AND SENTENCE CONTEXT EFFECTS

INTRODUCTION

Visual word recognition (lexical processing) provides the base for constructing meaning from text, as words are the primary meaning bearing element provided to the reader. Chapter 9 provides an extensive review of how people recognize words presented in isolation or in combination with a single-word partner. The evidence from isolated word recognition paradigms makes it clear that individual words contain a wealth of information of a variety of forms (e.g., visual features, orthography, phonology, morphology) and that people can utilize this information to recognize words rapidly and with relatively little conscious effort in the absence of sentence context. However, that is seldom the situation that we find ourselves in outside the laboratory. Far more often we find ourselves faced with recognizing words in the course of continuous silent reading. This is a task that may have word recognition at its base, but it is also a task in which the primary goal is comprehension of a larger meaningful message and as such is one that involves many additional processes that may not be engaged in act of recognizing a word that stands alone. This chapter addresses the issues related to lexical access during reading. Recognizing that the differential task demands of reading compared to isolated word recognition might affect the relative value of different sources of information (Balota, Paul, & Spieler, 1999), and given that reading places particular emphasis on processing for meaning, the first section examines the influence of lexical properties of words with particular emphasis on the meaning bearing properties of words. Although there is extensive evidence demonstrating that we are quite capable of recognizing words in the absence of context, there is also a large body of evidence demonstrating that context can influence word processing when it is present. The latter portion of the chapter takes up the issue of context effects on lexical access.

A. Lexical Properties

There are many different sources of information that are realized within the printed letter string of an individual word that might influence lexical access. These are factors that have figured prominently in studies of visual word recognition that have measured overt responses to words presented in isolation or in single word (priming) contexts and have been extensively reviewed in an earlier chapter of this volume (see Balota et al.). Many of these factors have similar effects whether one is faced with recognizing a word standing alone or one is reading for meaning. For example, there are clear effects of visual feature information in traditional word recognition paradigms and in studies of word recognition in the context of sentence reading. There is evidence of early phonological activation in reading from studies in which readers experience difficulty when they encounter a letter string with more than one possible pronunciation, for example, the letter string "wind" is pronounced differently when it refers to a weather condition than when it refers to a rotating action (e.g., Folk & Morris, 1995). Other studies have demonstrated that initial processing time on a word is affected by the existence of an unseen phonological partner, such as the pair "sale" and "sail" (Folk, 1999; Pollatsek, Lesch, Morris, & Rayner, 1996; Rayner, Pollatsek, & Binder, 1998), and there is evidence that these early phonological effects extend to scripts like Chinese that do not always make the script to sound relation transparent (see Pollatsek, Rayner, & Lee, 2000, for a review). However, the task demands of reading highlight the search for word meaning and the need to integrate that meaning with information gleaned from the text up to that point. And so, it is there that this review of lexical access in reading will begin. We will look at four aspects of word meaning and their respective roles in recognizing words in the course of reading for comprehension: morphology, word familiarity, word class, and lexical ambiguity.

1. Morphology

Morphemes are the smallest meaning bearing units of a word. Most long words in English are composed of more than one morpheme. One of the central questions in psycholinguistic research on morphology concerns the way in which this information is represented in the lexicon. That is, does each complex word have its own unique lexical entry (e.g., Fowler, Napps, & Feldman, 1985) or are they represented as a root with links to possible affixes (e.g., Taft & Forster, 1976). A second, related set of questions asks about how and when morphological units are identified and what role if any do they play in lexical access. Are morphemes active processing units in the recognition of morphologically complex words when people are engaged in continuous silent reading of connected text? Reading studies have exposed effects of morphological information in gaze duration on a word presented in sentence context. For example, Lima (1987) and Inhoff (1989a, 1989b) found differences in initial processing time between affixed and pseudoaffixed English words (e.g., relive and relish) and between compound and pseudocompound words (e.g., cowboy and carpet), respectively. In addition, there is evidence that initial fixation time on morphologically complex English words is influenced by the frequency of the morphological constituents that make up the word in addition to the frequency of the whole word form (Andrews, Miller, & Rayner, 2004; Niswander, Pollatsek, & Rayner, 2000) and that constituent frequency effects can be observed in cases in which the word form frequency is controlled (Juhasz, Starr, & Inhoff, 2003). Evidence of constituent frequency effects have also been documented in Finnish reading (Pollatsek, Hyona, & Bertram, 2000; Bertram & Hyona, 2003). These results suggest that morphological constituents are activated in the course of retrieving lexical representations.

So, how is a word decomposed into its morphological constituents prior to the word being recognized? There has been a recent spate of research reports (primarily masked priming studies) suggesting that morphological decomposition in English may be carried out in very early stages of word processing on the basis of orthographic information. Finnish readers spend less initial processing time on long compound words when vowel quality differs across the constituents that make up the word than when the vowel quality is consistent, suggesting that Finnish readers use vowel quality as a morphological segmentation cue (Bertram, Pollatsek, & Hyona, 2004).

Parafoveal preview manipulations have been used to ask questions about the time-course of morphological influence and the results have varied depending on the properties of the languages being investigated. To date the studies that have attempted to demonstrate preview benefit for morphological units during reading in English have failed (Lima, 1978; Inhoff, 1989a, 1989b; Kambe, 2004). In contrast, morphological preview benefits have been observed in Hebrew (e.g., Deutsch, Frost, Pelleg, Pollatsek, & Rayner, 2003; Deutsch, Frost, Pollatsek, & Rayner, 2005). These differences in the pattern of processing observed in English and Hebrew have been attributed to the morphological richness of Hebrew in comparison to English. However, it appears that the story may not be that simple. Recent preview experiments conducted in Finnish, another morphologically rich language, have failed to find evidence of a morphological preview benefit (Bertram & Hyona, in press).

2. Word Familiarity

Word frequency effects have been demonstrated in virtually every standard measure of word recognition, including naming, lexical decision, tachistoscopic report, semantic categorization, same–different judgments, initial reading as measured by fixation duration, and measures of the early time course of brain activity as reflected in the early components of event-related potentials (ERPs). The ubiquity of frequency effects has been taken as evidence that word frequency is a basic dimension of lexical processing, with more frequent words processed quickly compared with less frequent words.

Lexical familiarity, as assessed by printed word frequency, age of acquisition (AoA), or subjective familiarity rating influences a reader's initial processing time on a word, as measured by first fixation duration or gaze duration. Readers spend more time on less familiar than on more familiar words of equal length (e.g., Inhoff & Rayner, 1986; Juhasz & Rayner, 2003; Rayner & Duffy, 1986; Schmauder, Morris, & Poynor, 2000; Williams & Morris, 2004). These differences are observed even when word length, number of syllables, and word initial bigram and trigram frequency are controlled. High-frequency words are also more likely to be skipped than low-frequency words (e.g., Reichle, Pollatsek, Fisher, & Rayner, 1998). In addition to these initial processing time differences, the duration of the first fixation after the low-frequency word is often inflated compared to the high-frequency case. This is thought to reflect the processing of the low-frequency word spilling over onto the next fixation (Rayner, Sereno, Morris, Schmauder, & Clifton, 1990).

Gernsbacher (1984) suggested that printed word frequency may not be the best index of a reader's familiarity and experience with a word, given that people are exposed to words through spoken language as well as print. Juhasz and Rayner (2003) and Williams and Morris (2004) have demonstrated that other measures of word familiarity derived from subjective familiarity ratings, and AoA norms may represent unique (but overlapping) sources of variance to word reading time. Subjective familiarity is thought to be an index of frequency of exposure that is somewhat less biased to print exposure. Williams and Morris (2004) reported familiarity effects above and beyond what could be accounted for by printed word frequency norms. These effects were similar to the effects that have been observed in naming. Juhasz and Rayner (2003) reported unique contributions of AoA in the eye movement patterns of skilled readers and suggested that this measure may reflect differences in the quality of the semantic representation to a larger extent than the other two measures (see Juhasz, 2005, for more extensive treatment of these issues).

3. Word Class

One way of investigating the role of word meaning in lexical access is to compare reading behavior on words that differ in the degree to which they convey semantic content. For example, we can consider the extent to which linguistic distinctions between content and function words influence processing of those words in text. Content words denote entities, actions, and properties. They are derivational, have compositional meaning, and participate in productive compounding. In contrast, function words are defined by the grammatical relations or syntactic functions in which they participate. They have little if any lexical–semantic content do not participate in productive compounding or contribute to meaning in a compositional way. Because of these linguistic differences, some scholars have suggested that these two classes of words may be accessed differently.

A number of studies have demonstrated that when searching for a target letter in text passages, participants tend to show more detection errors (failure to notice the presence of the target letter) for a letter presented in a function word than for the same letter occurring in a content word. While the results are quite consistent, the theoretical interpretations are widely varied. Some theoretical accounts of this effect have emphasized differences at the level of lexical representation and process between the two word types (e.g., Healy, 1994), while others have emphasized differences beyond the level of lexical processing (e.g., Koriat & Greenberg, 1994). More recently, Greenberg, Healy, and Koriat (2004) have integrated their seemingly divergent views into a single model (Greenberg, Healy, & Koriat, 2004), incorporating both lexical level and text integration-level accounts of the missing letter effect.

Function and content word-processing differences have also been documented in more naturalistic reading tasks. Haberlandt and Graesser (1989) reported processing time differences between the two word classes in a self-paced reading task in which participants were simply told to read for comprehension. The later a word occurred in a sentence, the longer the processing time on that word, and the increase was greater for content words than for function words. This observation is consistent with the notion that there are observable processing differences between function and content words that occur beyond the lexical level.

Eye movement measures have revealed that readers are twice as likely to skip function words as they are to skip content words when reading technical prose (Carpenter & Just, 1983). However, differences in average word length and word frequency between the two word classes make it hard to know whether this result reflects differences in word class per se. Word skipping increases as word frequency increases (Reichle et al., 1998), even when word length and word class (nouns only) are controlled (Rayner, Sereno, & Raney, 1996). Word skipping is also known to increase as word length decreases (e.g., Rayner & McConkie, 1976) and as predictability increases (Balota, Pollatsek, & Rayner, 1985; Ehrlich & Rayner, 1981; Rayner & Well, 1996). Function words that are high frequency, short, and predictable are often skipped (Holmes & O'Regan, 1981; O'Regan, 1979; but also see O'Regan 1980, 1992). These findings clearly demonstrated effects very early in lexical processing, but it is not clear whether they should be attributed to differences between the two word classes per se. Schmauder et al. (2000) had participants read sentences that contained a critical function or content word and looked at processing time as well as word skipping measures. In a second experiment words from the two word classes were presented in a primed naming task. Neither the naming task, nor the initial processing time measures revealed evidence of lexical processing differences between function and content words when word length, frequency, and sentence position were controlled. However, there were interactions between word frequency and word class in later processing measures and these were taken to reflect the unique roles these two types of words play in constructing meaning from text.

In summary, the literature contrasting function and content word processing provides a great deal of evidence that differences in the grammatical functions of these two word types translate to processing differences. These differences are most clearly documented in tasks and measures thought to reflect text integration processes. In contrast, there is little evidence that lexical access is different for the two word classes.

4. Lexical Ambiguity

Lexically ambiguous words possess multiple meanings associated with a single orthographic form and, as such, they afford unique opportunities to examine the role of word meaning in lexical access. The questions addressed in the lexical ambiguity resolution literature have historically been seen as central to understanding the nature of the language-processing system more generally. In particular, much of this research has been dedicated to addressing the extent to which language processing is modular (Fodor, 1983) versus interactive (e.g., McClelland, 1987).

Early studies provided evidence that all meanings of an ambiguous word are activated, regardless of the context in which the word occurs (e.g., Frazier & Rayner, 1987; Rayner & Duffy, 1986; Swinney, 1979; Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982). Models based on this result came to be referred to as exhaustive access models and embraced the assumptions of modularity (Fodor, 1983): in particular, the assumption that there is an autonomous lexical processor that recognizes words on the basis of lexical properties alone, without benefit of meaning or context. Under this view all known meanings of an ambiguous word are accessed regardless of context, hence the label exhaustive access models. Other studies using similar methodologies provided evidence that given the appropriate context conditions, participants were faster to respond to a probe related to the context appropriate meaning of an ambiguous word than to a probe related to the context inappropriate meaning (Simpson & Kreuger, 1991; Tabossi, 1988; Tabossi, Colombo, & Job, 1987). The models developed to capture this result have been referred to as selective access models and are consistent with an interactive view of the language-processing system in which information derived from the context may interact with lexical information at the earliest stages of lexical processing. Under this view, given sufficiently constraining context, access may be limited to the context appropriate meaning of the ambiguous word.

Perhaps the most compelling evidence of exhaustive access came from cross-modal priming studies in which participants listened to a sentence or short passage and responded to a printed letter string presented visually. Presentation of the visual probe was calibrated to the auditory occurrence of an ambiguous word. Participants saw one of three possible probes. There was a related probe for each meaning of the ambiguous word and there was an unrelated probe. The time between probe and target could be manipulated to assess the status of word meanings over time. When the probe occurred within 200 ms of the target, participants were faster to respond to either of the meaning-related primes than to an un- related prime, suggesting that upon hearing the ambiguous word both meanings became available regardless of context (e.g., Onifer & Swinney, 1981; Seidenberg, et al., 1982; Swinney, 1979). As the time between the occurrence of the ambiguous word and the probe was lengthened, facilitation was limited to one meaning-related probe. Taken together, these findings suggested that readers initially access multiple meanings and rapidly select a single interpretation for incorporation into the ongoing text representation.

The timing of access also seems to depend on the relative frequency of the various meanings. In this chapter, we will use the term meaning dominance to refer to the extent to which one meaning is more likely to occur than another. The term balanced words refers to words with relatively equally likely interpretations, and biased words refers to words that have one interpretation that is much more likely than the other(s). Likelihood in these studies is typically operationally defined as the probability that a particular meaning is given as the first associative response to the word presented in isolation.

Meaning dominance effects observed in sentence priming studies demonstrated that for balanced ambiguous words the two meanings are activated close together in time (Seidenberg et al., 1982; Swinney, 1979). Biased words also showed evidence of multiple access (Onifer & Swinney, 1981), but the dominant meaning becomes available prior to the subordinate interpretation (Burgess & Simpson, 1988; Simpson, 1981; Simpson & Burgess, 1985), suggesting that access is frequency ordered.

Eye movement studies have also provided evidence of meaning dominance effects (Dopkins, Morris, & Rayner, 1992; Duffy, Morris, & Rayner, 1988; Rayner & Duffy, 1986; Rayner & Frazier, 1989; Rayner, Pacht,

& Duffy, 1994; Sereno, Pacht, & Rayner, 1992). In these studies, readers' eye movements were monitored as they read sentences, or short paragraphs and fixation time on the ambiguous word or a control word that was matched in length and frequency was measured. In general the findings from these studies can be summarized as follows. When neutral context precedes the ambiguous word, readers fixate longer on balanced ambiguous words than on biased words or on an unambiguous control word. However, when the reader encounters information that disambiguates to the subordinate interpretation following a biased ambiguous word, they spend more time than when disambiguating information follows a balanced ambiguous word or an unambiguous control word.

Duffy et al. (1988) also manipulated the sentence context preceding the ambiguous word. In this case, the effects of meaning dominance on initial processing time were reversed. Gaze duration on the balanced ambiguous word did not differ from the unambiguous control, suggesting that activation of the context appropriate interpretation exceeded that of the alternative interpretation, thereby reducing the competition between the two interpretations. In the case of the biased ambiguous words, the preceding context supported the subordinate interpretation of the word and readers looked longer at the ambiguous word than at an unambiguous control, suggesting that context had boosted activation of the weaker meaning so that it now competed with the otherwise dominant interpretation. Subsequent studies have yielded similar patterns of data with effects of similar magnitude (e.g., Dopkins et al., 1992; Folk & Morris, 1995; Rayner et al., 1994). The latter effect (longer processing time on a biased ambiguous word in context that supports the less likely interpretation) has been termed the subordinate bias effect or SBE (Rayner et al., 1994) and will be referred to as such throughout this chapter.

Casting questions of lexical ambiguity resolution as tests of modularity, the concept of exhaustive access was tied to the notion of an access process that was impervious to contextual influences, while the selective access accounts were decidedly context dependent. As this brief review has illustrated, there is now a large body of evidence demonstrating that readers access multiple meanings, and that access is influenced by the relative strength of the respective meanings of the word and the context in which it occurs. The picture that has emerged calls for hybrid models that account for rapid activation of multiple meanings and the early influence of sentence context.

There are a number of models of this sort in the literature (e.g., Duffy et al., 1988; Kawamoto, 1993; Simpson & Burgess, 1985; Twilley & Dixon, 2000; Kawamoto & Zemblidge, 1992). As an example, Duffy et al. (1988) proposed the re-ordered access model of lexical ambiguity resolution. This model was later instantiated in a connectionist architecture (Duffy, Kambe, & Rayner, 2001). The model makes four basic assumptions:

(1) access is exhaustive; (2) meaning dominance influences the relative activation of multiple meanings; (3) access is an interactive process in which disambiguating context preceding the ambiguous word may increase the activation of the context appropriate interpretation; and (4) the level of lexical activation of the context inappropriate meaning of the ambiguous word is unaffected by this process. While the re-ordered access model is representative of contemporary hybrid models, the various models differ in the extent to which they characterize exhaustive access as mandatory or highly likely and they differ in their view of the fate of the meaning that is not selected for integration into the discourse representation. Finally, although they all acknowledge that context may influence the order in which meanings become available, they say remarkably little about the properties of the sentence context that give rise to these effects. We will now deal with each of these issues in turn.

5. Selective Access Re-visited

The SBE has served as a test case for the assumption that access is exhaustive. Under the re-ordered access model the SBE has been taken as evidence that both meanings are activated, and that context may increase the activation of the subordinate interpretation to the point that the two meanings compete for selection. Alternative explanations of the SBE have been proposed by proponents of selective access models (e.g., Simpson & Kreuger, 1991; Kellas & Vu, 1999).

Under a strong selective access account the processing underlying the SBE is fundamentally different than that proposed by the re-ordered access model. According to this account, each meaning of an ambiguous word is stored separately in the lexicon and access is interactive. In the presence of supportive context, a single meaning may be activated without activating any other meanings associated with that form. The frequency of occurrence of the intended meaning of the word (not its frequency in relation to an alternative) and the context in which it occurs are the critical factors influencing initial processing time under this view. In contrast, the re-ordered access model assumes that all possible meanings of an ambiguous word share a common lexical entry. Context may influence the relative activation of candidate meanings, but contrary to the selective access view all meanings are accessed when the form is activated. Under this account, printed word form frequency, meaning dominance (the relative frequency of the possible meanings associated with that form), and the context in which the word occurs, are the critical factors, and exhaustive access is unavoidable.

Under the selective view the SBE is a word frequency effect. The subordinate interpretation of a biased word is in essence a low-frequency word, and so, when compared to a control word matched to the form frequency of the letter string (typically a much higher frequency than the subordinate meaning frequency), it takes longer to process. There is no competition between meanings. It is meaning frequency, not meaning dominance (the relative strength of the subordinate meaning with respect to the dominant meaning) that should determine processing time. Sereno et al. (1992) examined the SBE from this perspective. In their experiment readers encountered a biased ambiguous word in a sentence context that supported the subordinate interpretation. The frequency of the subordinate interpretation was estimated as the proportion of form frequency equivalent to the meaning bias. That is, an interpretation with a meaning with a bias of .15 was estimated to have a meaning frequency that was 15% of the form frequency of the ambiguous letter string. There were two control conditions in this experiment. One control was matched to the form frequency of the ambiguous word and the other was matched to the frequency of the context appropriate meaning. If access is selective, processing of a biased ambiguous word in a context that supports the subordinate interpretation and processing of an unambiguous word with a printed word frequency equivalent to the subordinate meaning frequency should not differ.

Sereno et al. (1992) obtained the typical SBE when the form frequency control condition was used as the comparison condition. Differences were also observed in three of the four processing measures that were reported when the meaning frequency control was used as the comparison condition. These results indicate that readers were not treating the subordinate interpretation of a biased ambiguous word like an unambiguous low-frequency word. The spillover and total time measures provide evidence of processing difficulty beyond that observed in the low-frequency unambiguous case, presumably due to the co-activation of the alternative interpretation of the ambiguous word (see also, Morris, in press; Sereno, O'Donnell, & Rayner, 2005). The fact that there are effects of meaning dominance above and beyond what is predicted by word frequency differences alone adds to the growing body of evidence that there is competition between simultaneously activated meanings.

Given the evidence that multiple meanings are accessed and compete for selection and the evidence that context can affect the status of that competition, one might ask about the limits of that contextual influence. Martin, Vu, Kellas, & Metcalf (1999) proposed that Duffy et al. (1988) and the numerous other eye movement studies that have demonstrated the SBE have lacked sufficient context strength to selectively access the subordinate meaning of an ambiguous word. Under this view exhaustive access and selective access are two extreme points on a common continuum, with selective access as an extreme case of contextual re-ordering. When the context is sufficiently constraining, only the context appropriate meaning of the ambiguous word should be accessed and no initial processing time cost should be observed. However, Binder and Rayner (1998) used Martin et al.'s materials (after eliminating some problematic items) and failed to replicate Martin et al.'s results whether they measured processing time in fixation duration or in self-paced reading time. Numerous other attempts to eliminate the SBE from the eye movement record of skilled readers by manipulating characteristics of the context have also failed (e.g., Binder, 2003; Dopkins et al., 1992; Kambe, Rayner, & Duffy, 2001; Morris & Binder, 2002; Rayner et al., 1994). Given this confluence of evidence, Binder and Rayner (1998) concluded that although context exerts an influence on the order and relative strength with which the possible meanings of an ambiguous word are accessed, even a strongly biasing subordinate context does not preclude activation of all candidate interpretations of a common word form. The evidence from eye movement monitoring studies is quite consistent on this point. But it is important to note that the data from sentence priming studies presents a somewhat different picture, with some studies showing evidence of selective access and others with seemingly similar manipulations not showing evidence consistent with selective access accounts and there appears to be no obvious way of categorizing these differences methodologically (Simpson, 1994; cf. Tabossi & Sbisa, 2001).

6. Fate of the Unselected Meaning

The research of the last 30 years has established that, with few exceptions, multiple meanings of an ambiguous word are activated at access, and a single meaning is rapidly selected. Within less than half a second the context inappropriate meanings of the word no longer show signs of activation. Models of ambiguity resolution have dealt with the change in the state of the unselected meaning in different ways. The recorded access model assumes that lexical activations of unselected meanings passively decay. In constrast, the activation-suppression model (Neill & Valdes, 1996) assumes that the unselected meaning is actively suppressed. If there is an active suppression mechanism at work, is it triggered by selection of the context appropriate interpretation (e.g., Binder & Morris, 1995; Gernsbacher, 1990; Gernsbacher & Faust, 1995; Gernsbacher, Robertson, & Werner, 2001; Morris & Binder, 2001; Simpson & Kang, 1994; Simpson & Adamopoulos, 2001), or do readers re-instate a prior episode and retrospectively inhibit the previously rejected interpretation when the situation calls for it (e.g., Neill, 1989; Neill & Valdes, 1996)? This is an area of active debate where many questions are yet to be answered, in fact there is an entire book dedicated to these issues (Gorfein, 2001).

B. Sentence Context Effects

Although there is still disagreement about the extent to which selective access is possible, and there is ongoing debate regarding the consequences of meaning selection on the status of the unselected meaning, there is agreement that context influences the status of the candidate meanings as they become available to the reader. There is also substantial evidence from research on the processing of unambiguous words indicating that readers are sensitive to contextual information. In eye movement studies these effects frequently emerge in first fixation and gaze duration, and this is consistent with the evidence from the lexical ambiguity resolution literature suggesting that context is influencing lexical access. One of the most consistent findings is that responses to words are faster when the word is preceded by a congruent context than when it is preceded by a neutral or incongruent context. For example, the word “treasure” is recognized more quickly in the sentence “The pirate found the treasure,” than in the sentence “The person liked the treasure,” or worse yet, “The house was destroyed by the treasure” (e.g., Balota et al., 1985; Ehrlich & Rayner, 1981; Fischler & Bloom, 1979, 1980; Foss, 1982; Hess, Foss, & Carroll, 1995; Schuberth, Spoehr, & Lane, 1981; Simpson, Casteel, & Burgess, 1989). The fact that context has an influence on word processing is clear, however, we are still far from reaching consensus on the processing mechanisms and/or the contextual factors that underlie these effects. The following section reviews the relevant evidence from studies examining the processing of ambiguous and unambiguous words encountered in the course of reading for comprehension.

1. Predictability

Some of the earliest studies of sentence context effects on word processing looked at the effect of predictability (Tulving & Gold, 1963; Tulving, Mandler, & Baumal, 1964; Morton, 1964). To predict is to declare or indicate in advance. In this section, the term “predictability” is used to refer to the extent readers might anticipate the identity of up- coming words based on the context in which they occur. This factor is typically operationalized either by measures of cloze probability in which participants are presented with a sentence fragment and asked to complete the sentence with the first word that comes to mind, or by rating tasks in which readers are asked to rate the likelihood that a sentence fragment would be continued with a particular word.

There are numerous studies showing that words that are predictable from the context in which they occur are processed more rapidly than words that are not predictable (e.g., Altarriba, Kroll, Sholl, & Rayner, 1996; Binder, Pollatsek, & Rayner, 1999; Ehrlich & Rayner, 1981; Frisson, Rayner, & Pickering, 2005; Inhoff, 1984; Lavigne, Vitu, & d’Ydewalle, 2000; Rayner, Ashby, Pollatsek, & Reichle, 2004; Rayner, & Well, 1996; Schilling, Rayner, & Chumbley, 1998; White, Rayner, & Liversedge, 2005; Zola, 1984). In addition, words are more likely to be skipped in a predictive context than in a neutral context (Altarriba et al., 1996; Balota et al., 1986; Dreighe, Brysbaert, Desment, & DeBaeke, 2004; Ehrlich & Rayner, 1981; Rayner et al., 2004) .

On the basis of the findings summarized in the previous two paragraphs one might conclude that readers are anticipating the upcoming word based on context and retrieving that lexical item prior to processing the perceptual input. However, we know that perceptual processing can be accomplished very rapidly. Reading is unimpaired when words are masked 100 ms after the start of each fixation and only slightly impaired when the exposure duration is reduced to 50 ms (e.g., Rayner, & Slowiaczek, 1981; Slowiaczek, & Rayner, 1987). These might be taken as estimates of the minimum amount of time required to extract the visual information from the page and mentally represent it in some sensory form. But even if we look at the typical fixation duration in reading standard text, readers spend just over a quarter of a second on any given word. In addition, we know that sentence contexts in natural language use are seldom sufficiently constraining to allow readers to accurately anticipate the next word in the discourse (Gough, Alford, & Holley-Wilcox, 1981) and assuming that predicting is a conscious strategy that readers apply to the process, erroneous predictions should be costly. Stanovich and West (1981, 1983) provided evidence that words with high cloze probabilities were named faster following a related sentence context than following a neutral or incongruent context. They reasoned that if readers were using the context to predict the next word prior to processing the perceptual input, then they should be slower to name the target in an incongruent context than a neutral context because they would have to reject their predicted completion before they could accurately name the target. Contrary to this hypothesis, they found no sign of inhibition when the word was named in an incongruent context.Taken together, these results suggest that although words that are highly likely to occur in a particular context enjoy a processing benefit, these effects cannot be fully accounted for by an anticipatory mechanism that predicts the upcoming word without benefit of perceptual input. Further evidence of this comes from studies in which words that are preceded by semantically related context are processed faster than words that are preceded by unrelated context, even when the word is not the predicted completion generated in a sentence completion task (Duffy, Henderson, & Morris, 1989; Morris, 1994; Morris & Folk, 1998). These results point to the need to postulate other mechanisms to account for sentence context effects.

2. Intralexical Priming

Lexical-level explanations of sentence context effects on word processing propose that contextual facilitation arises from word-to-word associations, or intralexical priming. According to this account, context effects are the result of activation spreading from related words in a context to the target word, speeding access to that word (e.g., Duffy et al., 1989; Fodor, 1983; Seidenberg et al., 1982), in much the same way that semantic-relatedness effects occur in word lists (e.g., Meyer & Schvaneveldt, 1976), or in the semantic priming experiments using single-word contexts that are reviewed in the visual word recognition chapter of this volume.

There is some evidence that when words that are semantically related occur very close to one another, the presence of the first word may prime the second. Sereno and Rayner (1992) demonstrated intralexical priming effects in sentence context using the fast-priming paradigm. This task paradigm combines properties of reading for comprehension with properties of traditional semantic priming paradigms minus the need for participants to make an overt response. Readers' eye movements are monitored as they read silently for comprehension. As the reader's eyes approach the target word, a random letter string occupies the target location. When the reader's eye lands on the target location, the letter string is replaced with a prime word for a brief period of time and then replaced with the target word. Sereno and Rayner found that readers spent less time on a target word presented in neutral sentence context when it was preceded by a semantically related prime presented for 30 ms than when it was preceded by an unrelated prime. There is also evidence of facilitation in gaze duration on a word when it is preceded by a close semantic associate (e.g., "king" preceded by "queen"), but only when the two words appear within a single clause (Carroll & Slowiaczek, 1986). There was no evidence of priming when a clause boundary was imposed between the two critical words. To the extent that there are intralexical priming effects in reading, they appear to be short lived.

Rayner et al. (1994) attempted to use intralexical priming to boost the activation of the subordinate interpretation of a biased ambiguous word so that it was accessed without competition from the dominant meaning. In one experiment the readers' first encounter with the subordinate sense of the ambiguous word was obtained through a paired associate task prior to a sentence reading task in which their eye movements were monitored. In a second experiment, both exposures to the word occurred within the context of a short passage. Gaze duration on the second encounter of the biased ambiguous word or an unambiguous control word (matched for length and frequency) was measured in each of these experiments. There was no evidence that priming by a lexical associate or explicit repetition of the biased ambiguous word in context that supported the subordinate interpretation at both encounters was sufficient to mediate the SBE (see also Morris & Binder, 2001).

While context effects on word processing in reading may arise in part from intralexical priming, the domain of influence of intralexical priming effects in continuous reading appears to be quite restricted. In addition, there is evidence that words are processed faster in related context than in an unrelated context, in the absence of lexical associates (e.g., Foss & Ross, 1983; Sharkey & Mitchell, 1985; Sharkey & Sharkey, 1992), and there is evidence that lexical relatedness alone is not always sufficient to produce this processing advantage (Duffy et al., 1989; Hess et al., 1995; Masson, 1986; Morris, 1994; O'Seaghdha, 1989; Potter, Moryadas, Abrams, & Noel, 1993; Simpson et al., 1989; Williams, 1988).

3. Interactive Sentence Context Effects

Conscious prediction mechanisms and automatic spreading activation between lexical-semantic associates are insufficient to account for the full range of contextual facilitation that has been observed. According to interactive accounts of contextual facilitation, emergent properties of the discourse representation may influence the processing of individual words during reading. In order to construct productive models of these effects, we need to take inventory of the properties of the discourse representation that might play a role, and those that do not.

Evidence that lexical relatedness alone is not sufficient to produce a word processing advantage exists in several different forms. Simpson et al. (1989) showed that sets of words that produced facilitation when embedded in a sentence context failed to produce the same effects when the same words were presented in a scrambled order. Conversely, Williams (1988) showed that a set of words that produced facilitation in the absence of a sentence frame did not yield facilitation when they were embedded in sentences.

Morris (1994) demonstrated that the time to read a word varied as a function of the sentence context, even when the words that made up that context did not change across conditions. In those experiments, participants read sentences like the following:

1. The waiter watched as the accountant balanced the ledger the second time.

2. The waiter who watched the accountant balanced the ledger the second time.

Participants' eye movements were monitored, and the primary dependent measure was gaze duration on the target word (ledger). The results of a sentence completion task were used to select target words that were produced as completions less than 15% of the time. Each sentence contained a verb, which was semantically related to the target ("balanced" in this example), and two nouns, each of which were related to the verb (waiter and accountant). One of the nouns, in conjunction with the verb, was highly related to the target word (accountant - balanced). The other noun, in conjunction with the verb, was related to a very different scenario (waiter - balanced). That is, when accountants balance, the sentence is about bookkeeping, and when waiters balance, the sentence is about restaurants. Control conditions were created by replacing the critical content words with neutral words (e.g., person, woman, saw, etc.). Readers spent less time on the target word (ledger) only when the accountant was balancing, and not when the sentence contained the same words, but the waiter was balancing. This effect was observed in both first fixation and gaze duration data. These results clearly implicate information beyond the lexical level in the word recognition process. However, there was also some evidence for intralexical priming, in this experiment. Processing time on the matrix verb was examined in some of the control conditions. In one set of controls, accountant or waiter appeared in conjunction with a neutral agent (e.g., man, woman, person) and in another set of controls, both potential subjects were represented by similarly neutral terms. Processing time on the verb was shorter following one semantically related noun and one neutral noun than following two neutral nouns. This processing time advantage was observed regardless of whether the semantically related noun was the agent of the action denoted by the verb or not, suggesting that it was the result of semantic association between the noun and verb, as opposed to properties of the discourse representation. This raises the possibility that discourse-level and intralexical-level accounts need not be mutually exclusive.

The experiments discussed thus far clearly demonstrate the need for an interactive account of sentence context effects and set some bounds on what does and does not interact. But they do not tell us much about how such interactive effects might occur. Schwanenflugel and LaCount (1988) defined contextual constraint in terms of semantic feature activation, and Tabossi has made a similar proposal regarding context effects on lexical ambiguity resolution. Schwanenflugel and LaCount proposed that high-constraint sentences impose greater feature restrictions on the spread of activation through the lexicon. Lexical access is facilitated to the extent that its semantic features are shared with the semantic features highlighted by the sentence context. If the sentence activates very general semantic features or a large number of specific features, then many words will be activated. This view makes the interesting prediction that words that are closely related to the predicted target word will only be activated to the extent that they also share semantic features with the context. So, for example, the word “shower,” processing of the word “shower” would not be facilitated in the sentence “The tired mother gave her dirty child a …,” even though it is closely related to the predicted completion, “bath” and it is a plausible completion for this sentence frame. Schwanenflugel and colleagues reported evidence of this in a series of cross-modal priming studies in which readers listened to a sentence and then made a lexical decision response to a visually presented probe. However, the responses in these experiments occurred more than a second after the offset of the sentence leaving open the possibility that multiple candidates were activated by the sentence context, and the best fit to the context was selected prior to making a response. Traxler and Foss (2000) showed that when an earlier processing measure is used, facilitation is observed for both predicted unambiguous sentence completions and for semantically related targets, contradicting the predictions of the feature activation account. Interestingly, although they found evidence that multiple lexical candidates were facilitated by the sentence context, they found no evidence of competition among the candidate completions.

The lack of evidence for competition in the Traxler and Foss studies is particularly interesting, given that the lexical ambiguity resolution literature provides numerous demonstrations of competition when multiple candidate interpretations are activated within a similar timeframe. The critical distinction between these two cases may be that in the unambiguous word experiments each candidate has a unique lexical entry (word form), while in the case of the ambiguous word there are two or more distinct semantic interpretations activated via a single word form. Perhaps the competition is specific to multiple candidates co-activated from a common form. Dopkins et al. (1992) took a slightly different approach to the question of the role of contextual constraint. They manipulated the type of disambiguating context that occurred prior to a biased ambiguous word. In one condition, the context supported the subordinate interpretation (positive evidence), as in “Having been heavily praised by the drinkers, the port was soon guzzled to the last drop.” In this case, the context prior to the ambiguous word “port” supports the subordinate beverage meaning of port through its association with drinking. This is representative of the disambiguating context that occurs most often in other published reports on ambiguity resolution. In another condition, the context biased toward the subordinate interpretation by ruling out the dominant interpretation (negative evidence) as in “Having been carried for miles by mule-train, the port was soon guzzled to the last drop.” In this case, the dominant “harbor” interpretation is ruled out since harbors cannot be carried by mule-train, leaving the subordinate “beverage” interpretation of port as the only viable option. Initial processing time (as measured by gaze duration) on the target word revealed that both negative and positive evidence were effective in boosting the activation of the subordinate interpretation to compete for selection. However, context that ruled out the dominant interpretation was no more effective than context that supported the subordinate interpretation.

4. Properties of the Discourse Representation

The majority of the work reviewed thus far has looked at contextual influences on lexical access occurring within a single sentence. We now consider some properties of the discourse representation that have been shown to influence comprehension and memory for text and ask to what extent these properties influence lexical access.

It has been suggested that entities related to the discourse topic are maintained in an active state and likewise that a shift in the topic of a discourse renders information related to the original topic less available for processing (e.g., Clifton & Ferreira, 1987; Gernsbacher, 1990; O'Brien, Duffy, & Myers, 1986). Binder and Morris (1995) examined the ability of discourse topic to influence lexical access by looking at the influence of topic on lexical ambiguity resolution. In the first experiment participants read short passages like the following that contained two instances of a balanced ambiguous word.

Meaning consistent. There was a lot of excitement at the bars downtown. Crowds of people were gathered outside the club (home) on the street. It appeared that someone had been hurt in the club that night. The police had been called.

Meaning switched. There was a lot of excitement at the bars downtown. Crowds of people were gathered outside the club (home) on the street. An hour earlier, a man was struck on the head with a club and robbed. The police had been called.

The context appropriate interpretation either remained consistent from first to second encounter or it was switched via information conveyed in the intervening context. Control conditions were created by replacing the first instance of the ambiguous word with an unambiguous control word that was matched for word length and word frequency with the ambiguous word. Initial processing time on the second occurrence of the balanced ambiguous word (henceforth the target word, in bold in the example) and on the context immediately following the ambiguity (the post-target region, underlined in the example) were measured. Looking at initial processing time on the target there was evidence that readers benefited from the repetition of the ambiguous word when the meaning remained consistent across encounters and there was no cost associated with the conditions in which the meaning switched.

Processing time in the post-target region showed the opposite pattern of results. There was a cost in initial processing time in the post-target region when the meaning switched and no benefit was observed in that region when the meaning was consistent with the first encounter. In another experiment, the meaning of the ambiguous word always switched from first to second encounter. But now, the intervening context either maintained the original topic of the passage or the topic shifted. A topic shift was operationalized as a change in the focal actor and a change in location. If the processing cost observed in the post-target region of the previous experiment is due to difficulty integrating the selected meaning into a discourse representation that includes the other meaning of that word, then that cost should diminish with a shift in discourse topic, as the previous concept is no longer in a highly active state in the discourse representation. The same logic applies to the initial processing time on the target word. If the benefit obtained in the meaning consistent condition emanates from the discourse representation, then it too should diminish with a shift in topic. Interestingly, the topic shift alleviated the processing difficulty in the post-target region in the inconsistent condition, but had no effect on the processing benefit observed on the target word in the consistent condition.

The initial processing time on the target word in Binder and Morris (1995) may not have been affected by the topic manipulation because accessing meaning is not affected by such global discourse factors or it could be because this particular topic manipulation was not effective. Several recent investigations of the role of linguistic focus in word processing in reading have provided results consistent with the first explanation. The focus of a sentence is said to be the most prominent or emphasized constituent in that sentence (Halliday, 1967). The focus of a sentence may be indicated through the use of wh-questions (e.g., Birch & Rayner, 1997; Blutner & Sommer, 1988; Cutler & Fodor, 1979), there-insertion sentences (e.g., Birch & Garnsey, 1995), or it-cleft sentence constructions (e.g., Birch & Garnsey, 1995; Birch & Rayner, 1997). Like topic, focusing is thought to enhance the relative availability of concepts in memory in spoken language comprehension (Gernsbacher & Jescheniak, 1995; Gernsbacher & Shroyer, 1989) and in reading (Birch & Garnsey, 1995; Birch & Rayner, 1997; Carpenter & Just, 1977; McKoon, Ward, Ratcliffe, & Sproat, 1993; Morris & Folk, 1998; Singer, 1976; Ward & Sturt, in press).

Blutner and Sommer (1988) manipulated focus through the use of wh-questions. They measured lexical decision time on a previously focused concept and found that focused concepts were responded to faster than non-focused concepts, suggesting that focus facilitates access. However, a recent series of eye movement studies paint a different picture. Birch and Rayner

(1997) failed to find initial processing effects for focused items while monitoring readers' eye movements in a silent reading task. In one experiment, Birch and Rayner syntactically directed focus through the use of it-cleft constructions, and in another experiment subjects read sentence pairs in which the first sentence was in the form of a wh-question that focused a particular entity in sentence two. In both cases they demonstrated differential effects of focus, but not in readers' initial processing time as measured by first fixation and gaze duration (see also, Morris & Folk, 1998; Ward & Sturt, in press). Although Morris and Folk (1998) also found no evidence of facilitation in initial processing time on the focused item, they observed a processing advantage for words that were semantically related to the focused item that occurred later in the clefted sentence, suggesting that the heightened prominence of the focused item in the discourse representation facilitated access of related words. These studies differed from Blutner and Sommer in that participants simply read the sentences as their eye movements were monitored. There was no overt response to the focused item required and the timeframe of processing captured in the first fixation and gaze duration measures of the eye movement record provide an earlier measure of processing than the lexical decision latency. Blutner and Sommer's results are consistent with the later processing advantage observed in the silent reading studies.

Several ambiguity studies have attempted to eliminate the SBE through manipulations of discourse-level variables (e.g., Binder, 2003; Kambe, Rayner, & Duffy, 2001; Morris & Binder, 2002; Wiley & Rayner, 2000). For example, in one experiment reported by Binder (2003), participants read passages that contained a biased ambiguous word. The sentence containing the ambiguous word (local context) supported the subordinate meaning and the discourse topic established in the first sentence of the passage (global context) was consistent, inconsistent, or neutral with the local context. Even when local sentence context and global discourse topic information converged in support of the subordinate interpretation, the SBE was not eliminated (see also Kambe et al., 2001; Morris & Binder, 2001). All of these studies obtained evidence that the SBE endured even when higher order discourse variables (global context, topic, conceptual repetition) and local sentence context were brought to bear.

Hopefully, this small sample of the work on context effects on lexical access has illustrated the progress that is being made toward understanding the limits of these effects. For example, the results reviewed here would suggest that the semantic content of the sentence context exerts a powerful influence on lexical access relative to the content of the more extended discourse (although we still have much to learn about "the semantic context"). In addition, although there is evidence that the message that emerges from successful syntactic parsing influences lexical processing, there is little evidence to suggest that syntactic relations per se have any direct influence on lexical activation. Finally, there is little evidence that linguistic devices known to increase the salience of concepts that have been retrieved from long-term memory also influence initial access to that information. Teasing apart those aspects of linguistic representation and process that have a direct influence on lexical access from those that do not is critical to developing more specific processing models.

Phoneme monitoring and lexical processing: Evidence for associative context effects

In this study, we investigated the role of associative semantic context in spoken word recognition by means of two different versions of the phoneme monitoring task. Using the standard phoneme monitoring procedure, in which subjects respond only to item-initial targets, we found no context effects. In Experiment 2, using the same stimuli but a modified procedure (the GPM), in which subjects respond to targets appearing anywhere in the word, we observed a clear context effect. Experiment 3 replicated these latter results and showed that the magnitude of this effect was unaffected by the number of associatively related items present in the experimental list.

We have shown in these experiments that the two versions of the phoneme monitoring task are differentially sensitive to lexical processing or, more specifically, to lexical context effects. 1 Two factors defining the GPM task favor a greater lexical contribution in the detection responses. These are the instruction set that defines for the subject where the targets to be detected are located (e.g., anywhere vs. initial position) and the actuallocation ofthe target phoneme in the target-bearing item. Let us tum first to the effect of the instruction set. Subjects performing in the standard (initial position) version can focus their attention upon the initial sounds of the target-bearing word and use this bottom-up information in making their response before lexical access has taken place. In contrast, subjects in the GPM task have no prior information about the position of the target phoneme and cannot attend to a specific part of the stimulus. As a consequence, they tend to rely more heavily upon the lexical code in order to detect target phonemes.

The role of attention in speech perception and word recognition has been investigated in the related area of phoneme restoration. The phoneme restoration effect (Warren, 1970) refers to a perceptual illusion in which listeners report hearing words intact even when a part of these words is replaced by noise. The strength of this perceptual illusion hasbeen shown to depend upon how much prior information subjects have about the position of the restored phoneme and about the lexical identity of the word (Nusbaum, Walley, Carrell, & Ressler, 1982; Samuel & Ressler, 1986). The more such information listeners have to help focus attention upon a specific part of a word, the less lexical knowledge influences their response. It has been argued (Samuel & Ressler, 1986) that although listeners normally direct their attention to the lexical level, under conditions such as those described here they are capable of focusing on the phonemic level, thereby limiting the influence of the lexicon. There is clear convergence between this interpretation and the one we have advanced. In both, it is claimed that when subjects are led by task demands to focus their attention on a specific part of the word, the effect of the lexicon is minimized.

The sensitivity of the GPM procedure to lexical context is due not only to the instruction set, but, as we have already mentioned, also to the sequential position of the target phoneme in the target-bearing word. The later the target phoneme arrives in the word, the further the lexical processing has progressed and the greater the effect of the lexicon.

Having shown the sensitivity of the GPM procedure to associative context effects, we attempted in Experiment 3 to specify more precisely the nature of these effects by manipulating the proportion of related items within the experimental list. Several authors (den Heyer et al., 1983, Tweedyet al., 1977) have taken an observed relationship between the proportion of related trials and the amount of contextual facilitation in lexical decision latencies to reflect subject strategies. As the proportion increases, so does the ability of subjects to consciously take advantage of the greater predictability of words.

To the extent that the pr~rtion manipulations in our experiment were adequate, the constant size of the context effect as a function of this manipulation suggests that our effect is not strategic.

Seidenberg et al. (1984) went further in their interpretation of proportionality effects and argued that these effects serve as one indicator of the locus of context effects. These authors compared the effect of varying proportions of related trials in two tasks-lexical decision and naming-and found proportionality effects only for the former. They concluded, on the basis of this result and convergent evidence from other experiments, that lexical decision latencies reflect some processes that take place after lexical access. These postIexical processes are argued to involve the integration of a word with its context or, in the case of lexical decisions, judgments of the subject about the relationship between the context and target words. PostIexical processes are opposed to prelexical processes, which are responsible for decoding the signal to recognize it as a particular word.

It is clearly essential to determine the extent to which the context effects observed with the GPM procedure have a prelexical and/or a postIexical processing locus. Unfortunately, simply showing that the subject has responded using the lexical representation does not allow us to identify the locus of this effect. Indeed, there is an important distinction to be made between the representation (pre-lexical or lexical) used in making the detection response and the stages of processing (prelexical or postIexical) where the context effects are localized. If the presentation of the prime facilitates access to the lexical representation of the target-bearing word, we can speak of a pre-lexical locus of these effects. Context effects have a postlexical locus, when, in contrast, it is not the recognition of the target-bearing word that is facilitated, but rather the decision processes taking place after recognition that involve judgments about the relationship between the prime and target-bearing word.

If the interpretation by Seidenberg et al. (1984) of proportionality effects is correct, our results provide some preliminary evidence that the context effects observed in the present study are not attributable to postIexical processes. It is premature on the sole basis of these results to draw general conclusions concerning the locus of these context effects. It is important to verify this interpretation by means of additional experimental findings.

Nonetheless, the results obtained in the present paper sug gest that the GPM task provides a promising way of studying lexical processing and evaluating context effects in spoken language understanding. The sensitivity of this task to processes underlying lexical access makes it useful in resolving the psycholinguist's dilemma of how to study the temporal properties of lexical processing.

Word-context Effects in Word Naming and Lexical Decision

The results of the present naming experiment as compared to our earlier lexical-decision study can be summarized in three points: (1) RTs with naming were, on average, 67 msec shorter than with lexical decision. (2) The facilitatory effect of a word context on the naming of subsequent context-related target words was smaller than that on lexical decisions to the same target words preceded by the same context words. More facilitation with lexical decision occurred already at 240-msec SOA, at which context-induced attentional processing seems to be hardly effective. (3) Whereas the earlier lexical-decision data had not shown a significant increase of the facilitatory effect over SOAs, with naming such increase does occur. However, the combined analysis including task as a factor does not allow the conclusion that the growth of facilitation over SOAs with naming is significantly larger than with lexical decision.

The first of these findings is in agreement with the literature (e.g. Forster and Chambers, 1973) and is compatible with the view that lexical decision requires time-consuming post-lexical processing (Forster, 1979, p. 29). As was set forth in the introduction, this property of a word-recognition task is one of the conditions that have to be fulfilled if meaning integration is to affect RTs. That meaning integration can indeed affect post-lexical processing in lexical decision is suggested by the second of the above results. Interpreting the differential facilitatory effects with naming and lexical decision along these lines, it appears that, overall, in our previous lexical-decision study meaning integration shortened the duration of the post-lexical stage of processing prime-related targets by 27msec. Finally, the finding that in the present naming study the facilitatory effect increased reliably over SOAs presumably has to be attributed to context-induced attentional processing becoming gradually more effective with increasing SOAs and evidences the occurrence of this process in word-context studies. In the introduction it was suggested that with lexical decision the increase of the facilitatory effect caused by context-induced attentional processing may have been overshadowed by the facilitation due to meaning integration; the increase of facilitation over SOAs with naming was predicted to occur since, as a consequence of switching from lexical decision to the naming task, meaning integration was thought to be rendered ineffective.

Unfortunately, the logical possibility of non-lexical naming (see Introduction) and the way in which we attempted to prevent that type of naming in the present study (namely, by presenting word targets only), allow alternative interpretations of findings 2 and 3, respectively: (2) If our attempt to prevent non-lexical naming was not altogether successful, the relics of that process may have had the same effect as rendering meaning integration ineffective by switching to the naming task, namely, less contextual facilitation in naming than in lexical decision. This is because facilitation can occur only on naming responses that come about via access of the prime-related target word's lexical representation. A similar dilution of facilitation cannot occur in lexical decision, since lexical access is a prerequisite in that task. Without having to regard meaning integration as a source of word-context effects on subsequent target word processing, the differential amount of facilitation in lexical decision and naming can thus be interpreted in terms of the "classical" two-process theory that only considers automatic spreading activation and context-induced attentional processing as contextual processes. Since context-induced attentional processing seems hardly effective at 240-msec SOA (Neely, 1977) and yet a 53-msec effect at this SOA with lexical decision is observed, this alternative interpretation requires the additional assumption that automatic spreading activation on its own can produce considerable facilitatory effects. (3) The facilitatory effect increased more over SOAs in the naming study than in the lexical-decision study, because the subjects in the former study (in which, due to the absence of nonword-target materials, the predictability of the target was larger than in the latter study) may have been engaged more in context-induced attentional processing. The present issue will only be definitely settled if the above pattern of results is found to replicate in future naming experiments in which logographic or highly irregularly spelled words constitute the experimental materials, since then, as in lexical decision, naming can only come about via lexical access.

If indeed such experiments also show the present differential effects of a related word context in lexical decision and naming, the above results may have considerable consequences for the intepretation of “priming” effects in lexical-decision studies. Very often the facilitatory and inhibitory context effects in these studies are well below 30 msec. And although it will often be inappropriate to compare context effects numerically across experiments using different materials? conditions and languages? the present data suggest that context effects of this size may not be true priming effects on word recognition, caused by context-induced attentional processing and/or automatic spreading activation. Instead, they may be caused entirely by the process of meaning integration shortening the duration of post-lexical processing in case of a related context-rarget word pair or of a congruous context-target sentence, and lengthening post-lexical processing in case of an unrelated context-target word pair or an incongruous context-target sentence.

In sum, the present data are compatible with the notion of an influence of meaning integration on the post-lexical processing duration in lexical decision, and the conclusion of Stanovich and West, that “If the goal of an investigation is to study sentence context effects on the process of word recognition, then the naming task is probably preferable (West and Stanovich, 1982, p. 385)”, may also apply to studies on the effects of a word context on subsequent word recognition.

Context effects in lexical access: A meta-analysis

The priming studies in this review showed a small effect of context on lexical access of about two tenths of a standard deviation: The appropriate interpretation of a word consistently showed greater priming than the inappropriate interpretation. Even this small ES is sufficient to disconfirm the modularity hypothesis. Moreover, despite claims by various researchers that certain variables influence access, the results of heterogeneity of variance tests suggested that the wide variation in ESs across studies may have been due simply to sampling error.

This kind of variation is to be expected, given the low power in these studies to detect small effects. To detect an effect of .2 at just greater than chance probability (.51) with an alpha of.05, the N of a study would have to be 140 (from Cohen's, 1977, power tables). None of the sample sizes in any of the studies reviewed was that high. The median sample size of an experiment was 35, and the largest sample in anyone experiment was 127. This means that, in half of the experiments in this review, the power to detect a small ES of.2 was, at best, just .13; in the experiment with the largest sample, the power to detect a small effect was just .38. One recommendation for future research, then, would be to increase the number of participants in an experiment, or at least to report ESs; this would facilitate comparisons with other studies, as compared with the cur rent practice of only making a categorical decision about significance.

A search for moderator variables gave no indication that modality of sentence presentation, length of context, and length of response influenced access. There was no evidence for ordered access or faster access for dominant interpretations in biasing contexts. However, there were relatively few studies in the review that manipulated meaning frequency, and fewer still that included ambiguous contexts in their manipulations. This indicates a need for more attention to frequency in future research.

There were indications that type of task and presentation point might influence the probability of finding a context effect. The use of the Stroop task as a measure of priming reduced the size of the appropriateness effect, which may indicate that acknowledged problems with the naming and lexical decision tasks (expectancy effects, backward priming, and the like) are contaminating the results of at least some of the studies. However, later studies were more aware of and, hence, more vigilant against potential contaminating factors in these tasks. Moreover, color naming has been used less frequently and has not been as carefully scrutinized as either of the other tasks, so its potential flaws may not yet be apparent. It is also possible that it is less sensitive to appropriateness effects.

Presenting the target prior to the end of the word seemed to reduce the size of the appropriateness effect. This suggests that a O-msec lSI might be too late to capture the initial stages of lexical access. However, because most of the studies using earlier presentation points were somewhat haphazard in their selection of presentation point, more use of the gating paradigm is needed to resolve this issue. Variations in target type seemed to be responsible for some of the variation in context effects in lexical access. Associates in our sample led to weaker context effects than did other types of targets, although they were still responsible for a weak appropriateness effect. This finding provides more evidence against modularity. According to Fodor (1983), the lexicon encodes only co-occurrence information among words; once a word is activated, linked lexical items should be obligatorily activated as well. An appropriateness effect for associates raises problems for this hypothesis.

Aside from the studies using associates, there were few semantic constraints on the type of targets chosen, so the targets were quite mixed in type. The loose restrictions on the type of targets used in many of the studies in this review may have contributed to the variation in the degree of context effects observed across studies. This lack of rigor, in turn, may reflect inadequate theories of word meaning; word meaning was ill defined in many of these papers. Targets were sometimes chosen solely by intuition but, more frequently, by just asking participants to generate features for words. More progress in the study of lexical access might be made by following the lead of Moss and Marslen-Wilson (1993), who used more principled distinctions among targets on the basis of core and peripheral aspects of word meaning. Although the distinctions they drew were based on a theory of semantic representation (cognitive economy in semantic networks) that is problematic (Ashcraft, 1976; Collins & Loftus, 1975), the attempt to move beyond intuition represents a step in the right direction.

The most dramatic claims for context effects has come from researchers who used constraining contexts, which are contexts that bias an interpretation by activating specific features of a noun's meaning. Although, as was discussed earlier, constraining contexts are not necessary to produce context effects (or even selective access), they do appear to be responsible for larger ESs than are undifferentiated contexts. The exceptions to this rule are the studies by Moss and Marslen-Wilson (1993) and Swinney (1991), which also used contexts that were designed to constrain the interpretation of the prime to one aspect of its meaning, but which failed to find significant effects of appropriateness. Swinney (1991) has suggested that this discrepancy is due partIy to Tabossi's materials and partly to the fact that her stimuli were in Italian. However, other studies (Kellas, Paul, Martin, & Simpson, 1991; Paul, 1992, and Paul et aI., 1992) that report selective access with constraining contexts use stimuli that are both different from Tabossi's and in English. In light of these discrepancies in results across studies, an account of what makes one researcher's constraining contexts different from another's is needed.

An obvious first place to look is in the procedures used to generate constraining contexts. Generally, all the constraining contexts were first generated by the experimenter; then, judges were asked to confirm the experimenter's intuitions that each sentence restricted the interpretation of the prime word to some attribute of its meaning. A priori constraints on the specific nature of these contexts were minimal. Experimenters avoided sentences that contained a strongly associated word that might prime the target even in the absence of the prime, as well as contexts that were strongly predictive of the prime itself, as opposed to an attribute of it. Also, as Swinney (1991) notes, in Tabossi's studies, judges were simply asked to use their intuitions and were given no theoretical guidance regarding what makes a context constraining; this was true for most of the studies in this category. These very broad constraints led to quite a range of contexts. For example, some contexts (those in Moss & Marslen-Wilson, 1993; Swinney, 1991; Tabossi, 1988a, 1988b; Tabossi et aI., 1987; Tabossi & Zardon, 1993) all seemed to make a particular feature salient through the use of weakly associated or related words. Alone, these words would not strongly suggest a particular feature, but together they made alternative features implausible. The example provided below illustrates this (the prime is the last word in the sentence and the target is the word in capital letters following it). Italicized words are words that are weakly associated or otherwise related to the target.

Context effects and the processing of spoken homophones

This study examined the role of context effects in the processing of spoken homophones in two experiments. In Experiment 1, Chinese speakers were presented with successively gated portions of a homophone in a sentence context, and they identified the homophone on the basis of its increasing acoustic information. The results indicate that sentence contexts influence the processing of Chinese homophones from early on, shortly after the acoustic onset of the word: when the homophone matches with sentence context, Chinese speakers can identify the appropriate meaning with less than half of the acoustic-phonetic information of the homophone. The results also indicate that lexical tonal information plays its role relatively late, usually at the onset of the vowel of a syllable, and that tonal information interacts with sentence context, leading to purely contextually driven interpretations of the lexical item. In Experiment 2, Chinese-English bilinguals were presented with a cross-language homophone in a sentence context, and they named a visual probe that had or did not have phonological overlaps with the homophone. The results show that prior sentence context significantly influences Chinese-English bilinguals' recognition of cross-language homophones, within the acoustic boundary of the word. Context helps bilingual listeners select the appropriate words at an early point when the acoustic signal is still ambiguous between Chinese and English and between various lexical candidates in the two languages.

Results from this study add new information on the operation of context effects in both monolingual and bilingual situations, and on the interaction between context and tonal information in homophone processing in Chinese. Consistent with our previous studies, our data support the context-dependency hypothesis that ambiguous meanings of a word may be selectively accessed from early on according to prior sentence context (Simpson 1981; Simpson & Krueger 1991; Tabossi 1988). In contrast, our study indicates that it is unlikely that Chinese speakers would exhaustively access all meanings of a homophone without using contextual information initially to constrain the access. Chinese speakers, faced with the extensive ambiguity created by massive homophones in the language, seem to have at their disposal a processing system that can rapidly disambiguate alternative homophone meanings during sentence comprehension. Such a processing system must be contextually driven early on to be able to operate efficiently.

There has been evidence in spoken word recognition that English speakers can identify a one-to-three syllable word in sentence context within about 200 msec, usually half or less of the acoustic signal of the word (Grosjean 1980; Marslen-Wilson 1987). According to Marslen-Wilson (1987), in English, there would be an average of 40 words still compatible with the available stimulus at 200 msec, when only the initial two phonemes are heard. In a bilingual situation, the problem may be even worse if lexical items are considered outside of context, because the number of lexical candidates compatible with 200 msec of a cross-language homophone may be even larger. Results from our study indicate that listeners can identify the correct meaning with only 33 percent of the homophone in the right context

(Experiment 1), and they can successfully respond to the target visual probe when only 150 msec of the auditory homophone was heard (Experiment 2). The 33 percent or 150 msec is insufficient acoustic information of a word. An examination of the acoustic waveforms of the 16 test words in Experiment 2 reveals that at 150 msec, all words except two included only the initial consonant plus some information of the vowel, at which point the word is still ambiguous between Chinese and English (and between various lexical candidates in each of the two languages). It is thus hard to imagine how speakers could recognize a word with only its minimal acoustic information, if they do not use contextual information from early on.

Tonal information in Chinese differentiates alternative meanings associated with the same syllable and thus reduces the potential number of homophones, although it does not eliminate homophony. We show in this study, however, that the role of tone in homophone processing is limited relative to the role of sentence context. Lexical tone can help the listener to disambiguate homophone meanings only when sufficient amount of the acoustic signal of the homophone is available, usually at the onset of the vowel in a syllable. Initially, only sentence context guides (or misguides) the word identification process. Later on, tonal information helps listeners to select among various candidates. In addition, tonal information does not always help. In some cases, listeners have detected the physical properties of the tone associated with the syllable, but context effects persist through the entire spectrum of the homophone, leading listeners to adhere to their incorrect identification. This pattern shows that context may initially override the physical properties of the lexical items during perception, leading to a garden-path of interpretation.

In short, results from the present study suggest that the successful recognition of spoken homophones depends on the interactions among the contextual, lexical, and phonological information in the sentence from early on. These results are best accounted for by interactive activation models of the sort in Kawamoto (1993), Marslen-Wilson (1987), McClelland (1987), and McClelland and Elman (1986). In these models, information processing flows both bottom-up and top-down, rather than strictly bottom-up, and lexical access and sentence context mutually influence one another at an early stage, rather than a stage at which context effects follow the completion of lexical access. These interactive models are largely inspired by or built on connectionist mechanisms that involve distributed representation, degrees of activation, and adaptation of connection strengths among processing units for phonological, lexical, syntactic, and semantic information of the sentence (Rumelhart, McClelland & the PDP Research Group, 1986).

In a connectionist perspective, the processing of spoken homophones can be viewed as an interactive process of constraint satisfaction: multiple sources of phonological, lexical, and contextual constraints either converge to facilitate the activation of relevant meanings, or compete to inhibit their activation. Thus, the product of processing at any stage is a result of the interactions among these sources of constraints, each of which may contribute different weights at a given time. Our goal in this line of research is to provide a comprehensive picture of the interactions among these various constraints, including context effects, homophone density effects, effects of lexical tones, and effects of the frequency of homophone meanings in the temporal course of the processing of spoken homophones.

Lexical Access during Sentence Comprehension (Re)Consideration of Context Effects

In all, the results from both of these experiments provide strong support for the conclusion that the access process for lexical items is isolable and autonomous at least with respect to effects of semantic context. That is, semantic contexts do not appear to direct lexical access, as was predicted by the Prior Decision Hypothesis. Thus, the access operation appears to be a stimulus (form)-driven process for which the entire inventory of information stored for a lexical form is made available to the sentence comprehension device. The results also support the existence of a post access decision process which acts to select a single meaning from those originally and momentarily accessed for involvement in further processing. This decision process apparently is completed at least by the time that three syllables of additional information have been processed (approximately 750-1000 milliseconds), even when no biasing context is present.

A few general comments concerning the posited post access decision process are in order. First, the normal time course of access, activation, and deactivation (for inappropriate meanings) in this process is clearly underestimated in this study. It is likely to be far less than the approximately 750-1000 milliseconds found in Experiment 2. Further as this decision process takes place within a 1000-millisecond period even for conditions containing no biasing context, one would expect it to be far faster in normal situations, where a context is typically present. Second, the nature of the decision process which chooses the relevant meaning of the ambiguity deserves some consideration. It may be that the process acts to suppress the level of activation of unchosen meanings. On the other hand, it may be that the single meaning which is chosen for an ambiguity is somehow made available to further (higher order) sentential processes in a manner which simply ignores the unchosen meanings. (For example, it could be that both meanings of the ambiguity are still somewhat activated following access.. but that the relevant meaning is shifted to what might be considered the "current" level of processing; presumably, it would be just this "current" level which can provide automatic semantic priming.) At present, there are no data which will allow us to directly choose between these quite different alternatives, and it is clear that further work on the nature of this decision process is in order.

Finally, because most words can, in fact, have different meanings (be these merely the different senses of a word or the totally different meanings comprising an unsystematic lexical ambiguity), it seems reasonable to suggest that the post access decision process posited here may be a general process. For any word, some subset of all the information which is originally accessed foi- that word may be selected for further processing and integration into ongoing sentential analysis. If so, only a single meaning for an ambiguous word, and only a single "sense" of an unambiguous word, would thus come to conscious awareness following this post access decision process. Semantic contexts apparently aid this selection process; the more the context restricts or determines the relevant sense of a word, the quicker the decision process will presumably take place. This model would fit with approaches taken by a number of authors (e.g., Collins & Loftus, 1975; Morton, 1969) on the access of semantic memory. It should be noted that while semantic contexts apparently do not affect access, there may be other types of information that will act upon the access phase of word recognition. Syntactic information, for example, may well serve to direct access in a way that semantic context cannot (see, e.g., Garrett, 1978; Fay, Note 6; Prather & Swinney, Note 7 ; Ryder, Note 8).

The model just sketched is, admittedly, underdetermined by the data. The nature of the claim being made is that sentence comprehension is not a totally interactive process; that is, that all kinds of information do not interact at all levels of processing. Certainly, it suggests that lexical access is basically a "bottom-up" or stimulus-driven process. This, however, is not at all to claim that this accessed information does not interact with other information. In fact, the data presented here could fit well with certain types of interactive models, such as that presented by Marslen-Wilson and his associates (e.g., Marslen-Wilson, 1975; Marslen-Wilson & Welsh, 1978), provided that certain constraints are placed on the interactions occurring around the access phase. In sum, however, these data appear to provide some evidence for autonomy of the lexical access process during sentence comprehension.

Lexical and Sentence Context Effects in Word Recognition

The present experiments replicated the very common finding that a sentence context leads to facilitation of the response to a subsequent word (e.g., Fischler & Bloom, 1979; Schwa-nenflugel & Shoben, 1985; Stanovich & West, 1979, 1981, 1983; West & Stanovich, 1982). In discussions contrasting autonomous and interactive views of word recognition, the existence and robustness of this effect have not been questioned, but its locus and the processes by which it operates have (e.g., Seidenberg, 1985; Tanenhaus & Lucas, 1987). In the following discussion, we shall attempt to evaluate the present results in light of these concerns.

The goal of the present research was to compare context effects when only lexical priming was possible to those arising from a normal sentence providing syntactic and message-level information. If word recognition is facilitated only by associations among words and not by concepts or propositions based on combinations of those words, then the facilitation should be independent of the order in which the words appear. It is this prediction that finds little support in the present studies. Normal sentences showed very large context effects, as they have in previous research, while scrambled versions of the same sentences showed only very modest effects. Normal sentences also showed larger context effects than the lexical primes showed in isolation in the pilot study. We cannot dismiss a contribution of lexical priming completely, however. The interaction of lag and relatedness in Experiment 3 must be attributed to the distance of the target from its lexical prime, regardless of sentence type. Also, an inspection of those near stimuli that had no words intervening between the lexical prime and target (0-lag stimuli) revealed that in Experiment 3, lexical priming in the scrambled condition did occur (50 ms, compared with 12 ms for all near trials). This effect for adjacent words in scrambled context was also found by Foss (1982). This post hoc comparison, however, is based on very few data points, in that only 6 of the 96 total sentences were of this 0-lag type, and these stimuli did not show greater priming than other near stimuli in either of the first two experiments (7 ms of inhibition in Experiment 1 and 28 ms of inhibition in Experiment 2). We conclude, therefore, that there are some residual effects of intralexical spreading activation, but they are inconsistent and usually quite small and cannot serve as a principal source of the larger context effects consistently found with intact sentences. In addition, the failure of lag and relatedness to interact with sentence type suggests an independence of lexical and sentence contributions to context effects. A further test of this independence would require a comparison of sentences with and without lexical primes but matched for the constraint (e.g., Cloze probability) that the context places on the target.

The difference obtained in the present experiments in the tendency of normal and scrambled contexts to speed word recognition is consistent with other studies using such stimuli (Foss, 1982; Masson, 1986; O'Seaghdha, 1989), and taken together, this research suggests that intralexical spreading activation by itself is a rather poor candidate to account for sentence context effects. Further support for this conclusion comes from research that has shown that spreading activation effects are relatively short lived and intolerant of disruption by intervening material (Dannenbring & Briand, 1982; Gough et al., 1981; R. Ratcliff, Hockley, & McKoon, 1985). Similar conclusions have been reached with respect to intralexical priming contributions to schema activation (Auble & Franks, 1983) and to the "plausibility effect" (J. E. Ratcliff, 1987), in which plausible sentences are processed more quickly than implausible sentences.

In conclusion, the principal result of the present set of experiments has been to show sentence-context effects on word recognition well beyond those attributable to intralexical spreading activation, and using a task that does not require a binary decision on the part of the subject. Although sentence-context effects have been shown many times, recent statements in favor of the lexical autonomy hypothesis have dismissed them either as intralexical effects (e.g., J. E. Ratcliff, 1987) or as postlexical effects of tasks requiring a discrete decision stage. Of course, the autonomy hypothesis (and the larger modularity view of which it is a part) does not rely solely on these arguments for its evidence, but these arguments have figured prominently in explanations of existing research on word recognition. The present results suggest that these explanations cannot account for all sentence context effects and that the case for syntactic and message-level contribution to word recognition should not yet be considered closed.

CONCLUSION

From the explanation above, it can be concluded that although there is extensive evidence demonstrating that we are quite capable of recognizing words in the absence of context, there is also a large body of evidence demonstrating that context can influence word processing when it is present. There are many different sources of information that are realized within the printed letter string of an individual word that might influence lexical access. Many of these factors have similar effects whether one is faced with recognizing a word standing alone or one is reading for meaning. For example, there are clear effects of visual feature information, phonological activation in reading from studies in which readers experience difficulty when they encounter a letter string with more than one possible pronunciation, initial processing time on a word is affected by the existence of an unseen phonological partner, phonological effects extend to scripts. There are four aspects of word meaning and their respective roles in recognizing words in the course of reading for comprehension, namely: morphology, word familiarity, word class, and lexical ambiguity.

Reading studies have exposed effects of morphological information in gaze duration on a word presented in sentence context. Differences in initial processing time between affixed and pseudoaffixed English words (e.g., relive and relish) and between compound and pseudocompound words (e.g., cowboy and carpet), respectively. In addition, there is evidence that initial fixation time on morphologically complex English words is influenced by the frequency of the morphological constituents that make up the word in addition to the frequency of the whole word form and that constituent frequency effects can be observed in cases in which the word form frequency is controlled. Morphological decomposition in English may be carried out in very early stages of word processing on the basis of orthographic information.

Lexical familiarity, as assessed by printed word frequency, age of acquisition (AoA), or subjective familiarity rating influences a reader's initial processing time on a word, as measured by first fixation duration or gaze duration. Readers spend more time on less familiar than on more familiar words of equal length and high-frequency words are also more likely to be skipped than low-frequency words. However printed word frequency may not be the best index of a reader's familiarity and experience with a word, given that people are exposed to words through spoken language as well as print. Other measures of word familiarity derived from subjective familiarity ratings, and AoA norms may represent unique (but overlapping) sources of variance to word reading time. Subjective familiarity is thought to be an index of frequency of exposure that is somewhat less biased to print exposure.

One way of investigating the role of word meaning in lexical access is to compare reading behavior on words that differ in the degree to which they convey semantic content. Content words denote entities, actions, and properties. They are derivational, have compositional meaning, and participate in productive compounding. In contrast, function words are defined by the grammatical relations or syntactic functions in which they participate. Participants tend to show more detection errors (failure to notice the presence of the target letter) for a letter presented in a function word than for the same letter occurring in a content word. In summary, the literature contrasting function and content word processing provides a great deal of evidence that differences in the grammatical functions of these two word types translate to processing differences. These differences are most clearly documented in tasks and measures thought to reflect text integration processes. In contrast, there is little evidence that lexical access is different for the two word classes.

Lexically ambiguous words possess multiple meanings associated with a single orthographic form and, as such, they afford unique opportunities to examine the role of word meaning in lexical access. All meanings of an ambiguous word are activated, regardless of the context in which the word occurs and participants were faster to respond to a probe related to the context appropriate meaning of an ambiguous word than to a probe related to the context inappropriate meaning. As the time between the occurrence of the ambiguous word and the probe was lengthened, facilitation was limited to one meaning-related probe. Taken together, suggested that readers initially access multiple meanings and rapidly select a single interpretation for incorporation into the ongoing text representation.

When neutral context precedes the ambiguous word, readers fixate longer on balanced ambiguous words than on biased words or on an unambiguous control word. However, when the reader encounters information that disambiguates to the subordinate interpretation following a biased ambiguous word, they spend more time than when disambiguating information follows a balanced ambiguous word or an unambiguous control word.

Context may influence the relative activation of candidate meanings, but contrary to the selective access view all meanings are accessed when the form is activated. Under this account, printed word form frequency, meaning dominance (the relative frequency of the possible meanings associated with that form), and the context in which the word occurs, are the critical factors, and exhaustive access is unavoidable. Under the selective view the SBE is a word frequency effect. The subordinate interpretation of a biased word is in essence a low-frequency word, and so, when compared to a control word matched to the form frequency of the letter string (typically a much higher frequency than the subordinate meaning frequency), it takes longer to process. There is no competition between meanings. It is meaning frequency, not meaning dominance (the relative strength of the subordinate meaning with respect to the dominant meaning) that should determine processing time. Readers encountered a biased ambiguous word in a sentence context that supported the subordinate interpretation. The frequency of the subordinate interpretation was estimated as the proportion of form frequency equivalent to the meaning bias. There were two control conditions in this experiment. One control was matched to the form frequency of the ambiguous word and the other was matched to the frequency of the context appropriate meaning. If access is selective, processing of a biased ambiguous word in a context that supports the subordinate interpretation and processing of an unambiguous word with a printed word frequency equivalent to the subordinate meaning frequency should not differ.

Multiple meanings of an ambiguous word are activated at access, and a single meaning is rapidly selected. Within less than half a second the context inappropriate meanings of the word no longer show signs of activation. Models of ambiguity resolution have dealt with the change in the state of the unselected meaning in different ways. The recorded access model assumes that lexical activations of unselected meanings passively decay. In constrast, the activation-suppression model assumes that the unselected meaning is actively suppressed. If there is an active suppression mechanism at work, is it triggered by selection of the context appropriate interpretation or do readers re-instate a prior episode and retrospectively inhibit the previously rejected interpretation when the situation calls for it. This is an area of active debate where many questions are yet to be answered, in fact there is an entire book dedicated to these issues.

One of the most consistent findings is that responses to words are faster when the word is preceded by a congruent context than when it is preceded by a neutral or incongruent context. The following section reviews the relevant evidence from studies examining the processing of ambiguous and unambiguous words encountered in the course of reading for comprehension.

CHAPTER 9
SEMANTIC MEMORY

INTRODUCTION

During the decade ending around 1972, Quillian and colleagues introduced a seminal computational model that they called semantic memory. This was less than a century after memory itself was introduced to modern psychology by, e.g., Wundt, Ebbinghaus, and James in the late 1800s, and semantics per se was brought into existence as a field of study within linguistics by Bréal. Quillian's semantic memory continues to influence psychology, psycholinguistics, and cognitive neuroscience. The PsycInfo database con- tains over 2400 publications containing the keyword "semantic memory," from 1966 (when Quillian's dissertation was completed) to 2005. The most important paradigms involving semantic memory today are the "neuro" paradigms, which involve the relationship of semantic memory to brain structure or function. Figure 1 illustrates the transition of semantic memory research from an almost purely normal-literature phe- nomenon to one increasingly dominated by neuro paradigms. Over half (54%) of all semantic memory publications have been neuro-related; for the past decade, almost 3/4 (72%) have been neuro-related.

What do psychologists mean when they use the term semantic memory? Almost half a century ago, in 1972, Endel Tulving suggested partitioning the human long-term memory system into two distinct stores: an episodic store that contains memories that are linked to a particular time and place (e.g., where you parked your bicycle this morning), and a semantic store that contains more general knowledge about the world (e.g., what a bicycle looks like, or is used for). Tulving's proposal was widely adopted, and now many psychologists and cognitive neuroscientists consider episodic and semantic memory to be components of the declarative (or explicit) branch of the long-term memory system. Motor knowledge about how you actually ride a bicycle, in contrast, is generally described as a procedural skill that is part of another branch of long-term memory—the nondeclarative, or implicit, memory system. This system encompasses knowledge to which we do not have conscious access, but that nevertheless affects our behavior (Squire, 1987).

Scientists who study semantic memory are usually concerned with word meaning and how the meanings of multiple words can be combined to understand longer text segments. They have also been concerned with how people acquire word meanings, how they use them to draw appropriate inferences, how they can efficiently store and search vast amounts of information, why some systems of categorization seem better and more natural than others, what compo-nents of meaning become active immedi-ately when we encounter words, and why various kinds of brain damage can lead to specific patterns of loss of word meanings. Most research has been conducted using tasks that present simple statements (e.g., “A dog is an animal.”) or combinations of words in temporal sequence (e.g., dog -> cat) and ask people to make judgments about the stimuli.

A. Quillian’s semantic memory

Quillian’s semantic memory was first a theory of human long-term memory, and second a series of computer simulations of certain types of language processing. This section contains our interpretative summary of semantic memory theory as reflected in Quillian’s publications over a period of approximately a decade (Quillian, 1961, 1962, 1966, 1967, 1968, 1969; Collins & Quillian, 1972). Our goal in this section is to present an accurate picture of Quillian’s framework, including certain elements that are not widely recognized today.1

During the late 1950s and early 1960s, it was widely conceded that lexical ambiguity was the greatest problem facing syntax-based natural language-recognition systems. The most promising of the potential solutions considered at the time was the “thesaurus method,” which involved looking up each word of the sentence in an online thesaurus, performing set intersections on the resulting lists of words, and using the intersections to resolve ambiguities in the words of the sentence (e.g., Masterman, 1957). Quillian’s model substantially extended the thesaurus method: he used semantic relations among the words in a text either to represent an “understanding” of the text virtually without syntax (earlier versions), or via parallel semantic and syntactic analysis (later versions). It was because of the need to capture the semantic relations of the English lexicon that the structures and processes of semantic memory were conceived. Because of the vari- able terminology used by Quillian and subsequent writers, and to avoid irrelevant theo- retical connotations, our exposition will use the more neutral word entry in place of token, concept, word, property, unit node, frame, and so on, and binding in place of con- nection, link, type, feature, and so on. The intended metaphor is that of entries in a dictionary or encyclopedia bound together into intricate constellations of meaning as they recur in the bodies of definitions.

1. **Structures**

   1) *Entries and bindings*

A Quillian semantic memory consisted of a set of entries, interconnected with arbitrarily complex bindings. Each entry corresponded to a conceptual notion, including but not limited to things like words and propositions. Each entry had associated with it a set of proximate bindings. All bindings had a unidirectional pointer to a predicate (attribute) and some number (usually zero or one) of unidirectional pointers to values. Note that the attribute was in effect a label on the binding, an idea probably influenced by Quillian's teachers, Simon and Newell, who in turn had been influenced by the work of the German psychologist Otto Selz (Simon, 1981). The predicates and values in bindings were all entries in the semantic memory themselves. For example, an entry for canary might have a binding with the attribute label color and the value yellow. An English-like approximation would be "The color of canary is yellow." In turn, the structure of attributes and values of bindings could have bindings themselves, for example, the attribute label color in the example might have a nested binding with the attribute location and the value body-surface, the value yellow might have a nested binding with the attribute color-saturation and the value pale: "The color of the body-surface of canary is yellow with a color-saturation of pale." In addition to subnetworks of attributes and values, bindings also optionally had weights that indicated their physical strength, or intensity, as well as values indicating their importance to the entry of which they were a part in terms of number and criteriality, notions borrowed from the classical categorization theory of Bruner, Goodnow, and Austin (1956).

Quillian emphasized several structural aspects of this view of semantic memory. First, each attribute and value of each binding was itself an entry, with its own set of proximate bindings. Second, there was no strict hierarchy in a semantic memory as a whole. Each entry was the root of its own hierarchy, but simultaneously was subsumed under an unlimited number of other hierarchies branching from other entries. Third, the chains of bindings leading away from a given root entry could refer back recursively to the root entry itself; that is, the entry canary could appear as an attribute or value of an indirect binding under the root entry canary. For example, starting with the verb to fly, we might find the following loop: fly, wing, feather, bird, fly. Similar longer or shorter recursive loops are quite common.

2) Superset bindings and cognitive economy

There was one kind of binding which was particularly important in Quillian's model, the superset binding, also known as the isa binding. Superset bindings indicated set inclusion, so for example, canary would have a binding with the attribute superset and the value finch (a canary is a type of finch). Quillian's model made substantial use of the transitivity of the superset relationship to implement the principle of cognitive economy. That is, if bird was a superset of finch, and finch a superset of canary, there was no a priori need for a superset binding corresponding to bird in the canary entry, since this relationship could be inferred using transitivity.2 More than one superset binding could be present; for example, the entry for canary might have superset bindings for both finch and for bird. Multiple superset bindings were motivated by experimental results show- ing that high frequency or highly criterial attributes that were logically indirect (e.g., bird) could be accessed more quickly than less common, less criterial attributes that were logically more direct (e.g., finch; Collins & Quillian, 1972).

2. Processes

1) Spreading activation

One can intuitively visualize the very large constellation of entries and bindings in a semantic memory as being arranged in three-dimensional space such that the distance between entries is a function of how many stages of binding separate them.3 When an entry is accessed, activation spreads out along its bindings, passing through connected entries and in turn out along their bindings. Quillian (1966, p. 72) expressed this vividly as "firing an activation sphere," evoking an imaginary bubble expanding out from the original entry. The speed and priority of this breadth-first search could in principle be modulated by number (i.e., quantity) and criteriality tags in the bindings (although the extent to which Quillian applied this feature is not clear). The search would terminate after a certain number of entries had been examined. As activation reached an entry,   the entry was given a temporary marker, or activation tag, which identified the origin of the current activation sphere and also contained a pointer back to the immediately pre- ceding entry or binding. If a new entry reached by a thread of this activation sphere already had an activation tag from the sphere's origin, the thread ended there.

2) *Intersections, paths*

As understanding of a text progressed, spheres of activation were rapidly fired from successive words (or other grammatical units) so that multiple parallel spheres were active at the same time. In this case, when activation reached an entry that had a tag from a different sphere of activation, the bidirectional path between the initial entries specified an intersection, corresponding to a possibly complex inference involving the entries. For example, in the sentence John put the canary into a cage, the binding of canary as a bird, and the binding of bird to pet (birds can be pets), plus the binding of pet and bird to cage (pet birds are kept in cages), would lead to the inference that the canary in the sentence could be a pet.

Since spreading activation as described by Quillian was an automatic process not influenced by syntax or other context, each new intersection was evaluated based on its context, including the surrounding syntax, but also in terms of form tests specific to the entries. Evaluations of intersections could be consciousan "interrupt" occurred, in Quillian's computational metaphor. For example, in the sentence He saw a canary fly by the cage, presumably the same intersection involving canary–(bird–pet)–cage would be found, but the inference that the canary could be a pet would not be activated, be- cause it is inconsistent with the syntactic frame. If the evaluation of an intersection failed, then it was abandoned and the spreading activation/evaluation process continued until the processing limit was exceeded. As validated intersections were found, they were unified with previously activated intersections to create a situation-specific representation, in the same format as existing entries, which could become a new entry in semantic memory, for example, an entry corresponding to a text or to an episode.4 There is a fairly good fit between Quillian's notion of the conscious evaluation of intersections leading to the construction of situation-specific conceptual structures, with Chafe's (1994) focal/active, peripheral/semi-active, and unconscious/inactive states of knowledge during discourse, with short intonation units corresponding to brief activations of ideas, which then quickly recede to semi-activation. This basic concept also can be aligned with Barsalou's (1999) idea that network-like conceptual representations are created dynamically and situation-dependently, thereby creating performance differ- ences in various categorization tasks. For Barsalou, the dynamic structures are created from modality-specific perceptual codes rather than amodal structures, but there is nothing fundamentally incompatible between the formulations of Barsalou and Quillian, if the perceptual codes were organized, accessed, and activated as a semantic memory; in fact, the goal of representing perceptual processing and knowledge was stated several times by Quillian (1966, 1968). These commonalities in the thinking of Quillian, Chafe, and Barsalou suggest that, as Chafe put it, "In the long run, it may be less fruitful to think of something being in memory, or retrieving something from memory, than to view these phenomena in terms of activation" (p. 53, italics in original).

3) Supersets, identicalness

As newly activated intersections were verified and unified with previous material, it was necessary to determine which constituent entities were identical to each other, or in Quillian's terminology, whether two entities could be identified. For example, one way to understand the sentence John bought a canary, but the bird died is to infer that canary and bird can refer to the same, identical entity. Quillian's principal identity-verification strategy was based on the superset intersection, an intersection that consisted only of superset bindings. Note that if one entry was directly above another entry (canary–bird), then in most cases they could stand in the identity relationship. On the other hand, if they shared a common superset but one was not the superset of the other (canary–pigeon), then they may or may not be identical, depending on the context and the values of bindings with shared attributes. Much of the early reaction-time (RT) literature on semantic memory by Collins, Quillian, and others, was based on verifying sentences like A canary is a bird, A canary is a robin, and so on (see Collins & Quillian, 1972, for a discussion and overview). As a result of these influential experiments, and of discussions that emphasized superset intersections more than intersections involving other types of bindings, the Quillian semantic memory has sometimes been characterized mistakenly as if it had only superset bindings. In addition to descriptive bindings (e.g., color, habitat), there were several other bindings that could be used in logical inference in various ways (including probabilistically): similarity, part, proximity, adjacency, consequence, precedence, and parent. All of these bindings, and an unlimited number of others, were part of the basic structure of a semantic memory and were traversed by spreading activation.

B. Generic Semantic Memory

Quillian's conception of a semantic memory was both a set of theoretical assumptions or hypotheses about human memory and language, and an evolving specification for computational simulations based on them. The two simulations implemented by Quillian, of finding meaningful intersections between specific words, and of building up mean- ingful representations of English text, were illustrative, but did not exhaust the intended range of applications of the theory, which reached beyond the domain of language un- derstanding. For example, "spatio-visual" memory, the recognition of objects through sensory perception, the generation and storage of imagery, the induction and learning of conceptual knowledge, and analogy and metaphor were all considered prime candidates for modeling via a semantic memory, at least insofar as "the same static store of infor- mation" could underlie such other mental phenomena, "rather than supposing that these rely on separate memory structures, even though, such a memory would then have to be richer in interlinkages than that we shall utilize here" (Quillian, 1966, p. 22; see also Collins & Quillian, 1972; Collins, 1975).

During the decade in which the Quillian semantic memory model was being actively developed, there were several extensions and modifications reflecting the results of psy- chological experiments. Since Quillian's final publication on semantic memory (Collins & Quillian, 1972), it has been the basis, as was shown in Figure 1, of a continuously growing body of work. Several subsequent avenues of research were connected directly to the Quillian model, for example, its application to the problems of programmed learn- ing and human reasoning (Carbonell & Collins, 1973; Collins, 1978); in addition, several computational models which were quite similar to the

Quillian framework (e.g., Schank, 1975; Rumelhart, Lindsay, & Norman, 1972; Anderson & Bower, 1973) were developed during the same period. However, most subsequent work was based on particular com- ponents of Quillian's semantic memory, often altered in meaning, scope, or form. This section contains a brief survey of the key influences on the "genericization" of semantic memory.

1. Consciousness

While Quillian discussed the interaction of semantic memory and consciousness, he did not emphasize it. Recall that the basic process of spreading activation was uncon- scious, but when an intersection was found, the subsequent evaluation of the intersec- tion could involve some degree of consciousness. Tulving (1972) noticed that work stemming from or related to Quillian's framework, while undeniably related to human memory, was quite different from the list-learning or paired-associate learning para- digms that had dominated human memory research up through the 1960s (but cf. Tulving, 1962; Bower, 1972). He proposed two qualitatively different kinds of human memory, one based on episodes, the other on knowledge. He delineated several differ- ences between them, and he called them *episodic* and *semantic* memory. Tulving (1985) fractionated long-term memory further, adding *procedural* memory (but cf. Squire, 1994; Schacter & Tulving, 1994).

Tulving (1985) identified episodic, semantic, and procedural memory with three kinds of conscious experience: *autonoetic, noetic,* and *anoetic,* respectively. The three types of memory and their alternate conceptions as types of consciousness were hierarchical: to say that a certain memory event was episodic, for example, was to say that it was associated with autonoetic consciousness (and possibly with noetic and anoetic con- sciousness); semantic memory events were associated with noetic consciousness (and possibly anoetic, but not, presumably, autonoetic consciousness). One result of Tulving's proposal has been an equivocation on "semantic memory." Under Tulving's formulation, the dynamics of the system were secondary; semantic memory was defined in contrast to other memory systems in terms of the associated personal, conscious experience. This conflicts directly with central elements of Quillian's framework: constellations of entries bound together, and the twin processes of unconscious spreading activation followed by conscious evaluation of intersections. This equivocation has resulted in some confusion in the literature (see Tulving, 1983, for further discussion). In summary, Tulving con- ceives of semantic memory as a subset of what Quillian called semantic memory, prima- rily in contrast to episodic memory; most versions of generic semantic memory adopt Tulving's semantic versus episodic distinction.

2. Spreading Activation

Recall that in Quillian's framework, spreading activation involved a complex activa- tion marker spread among entries via their bindings. The marker contained information used to control the further spread of activation and to reconstruct the full paths of inter- sections. The entries and bindings in the resulting path were unified with the develop- ing overall conceptual representation of a text, and could result in the creation of new conceptual entries. Collins and Loftus (1975) redefined spreading activation as a continuously variable process in which activation spread in a fluid-like manner through bindings whose capacity and length were a function of relatedness, criteriality, and so on. The amount of available activation was limited, and it drained away from reservoir- like entries over time. An entry was triggered into activity when it reached a threshold level of activation, and then it began to spread activation through its outputs. Because of the emphasis on bindings of different length and capacity, and on the build-up and fad- ing away of activation, this kind of model was well-suited to describe temporal phe- nomena such as semantic priming (section 4.2). However, since the concept of a discrete path between entries was abandoned in favor of the activation of individual entries, new concepts could no longer be derived from meaningful, context-specific intersections of existing ones.

In a further fractionation of semantic memory, Collins and Loftus's (1975) revision also assumed two largely independent, specialized semantic memories, one for the lexicon, the other for nonlexical knowledge. This was based on the observation that priming can be found independently for similar-sounding words, conceptually related words, or both (but Collins, 1975, continued to have reservations about this–and any other–division of semantic memory). The distinction between lexical versus semantic memory is fairly common in the literature today; generic semantic memory usually but not always features at least one independent lexicon (see Coltheart, 2004, for a discus- sion). See Schank (1976) for yet another proposed apportionment of lexical, episodic, and other elements of semantic memory.

3. Clinical Assessment

In seminal work, Warrington (1975) applied certain aspects of Quillian's framework to clinical memory disorders. In her work, the superset binding hierarchy was empha- sized over the connectivity among other types of bindings and the processes of spread- ing activation and evaluation, a tendency that has characterized neuropsychologists' subsequent interest in semantic memory. While a range of different methods have been used to test semantic memory in clinical settings, the tendency has been to examine the standard repertory of neuropsychological tests of language function, and to choose those that include tests of concept/word knowledge but do not emphasize syntax, phonology, communicative competence, personal memory, and so on. As a result of this selection process, clinical tests of semantic memory have been based to a large extent on two categories of tests: confrontation naming of pictures (usually the Boston Naming Test; Kaplan, Goodglass, & Weintraub, 1983), and word fluency, or "con- trolled word association," tests in which subjects must generate words that belong to a certain category, such as items found in supermarkets, or that begin with certain letters (FAS). Note that the category versus lexical distinction in fluency tests corresponds somewhat to Collins and Loftus's (1975) distinction between lexical and nonlexical se- mantic memory.

It has been pointed out that these tests examine only a small subset of semantic mem- ory capabilities, and that they involve cognitive processes (such as attention, working memory, and strategy deployment) not specific to semantic memory (e.g., Rende, Ramsberger, & Miyake, 2002; Ober, 2002; Shenaut & Ober, 1996). In response to the limitations of naming and fluency in the assessment of semantic memory, more nuanced test batteries have been created, such as that of Hodges, Salmon, and Butters (1992), which was based on using the same set of 48 test items in naming, category fluency, sort- ing, picture–word matching, and verbal generation of definitions. Also, some alternative clinical instruments involving semantic memory have been developed, such as the Pyramids and Palm Trees test (Howard & Patterson, 1992). Still, as the importance of clinically based studies has increased in the semantic memory literature, clinical investigators have contributed to the process of genericization by frequently altering or simplifying the underlying theory and by relying on just a few limited and/or insuffi- ciently specific neuropsychological tests.

C. Theoretical Extensions, Consequences, and Divergences

The decade during which Quillian was developing semantic memory theory ushered in a period of very active theoretical development in psychology and linguistics, and so it can be difficult to determine the original source of contemporary ideas. However, there is a group of important theoretical advances that are allied to Quillian's frame- work, even though some may not have cited Quillian's work explicitly, or may have referred to a more generic conception of semantic memory. This section briefly surveys a selection of them.

1. Eco's Model Q

The semiotician Umberto Eco (1976) pointed out that meaning cannot adequately be represented in terms of decomposition into sets of elements or features, because in order for the features to have meaning within the system, they in turn would need to be de- composed into further sets of elements and features, and so on, in what he called infinite semantic recursivity. His solution to this problem was what he called "Model Q" (the "Q" is for "Quillian"), basically semantic memory à la Quillian (1968). Model Q's main ad- vantage for Eco was that in contrast to other approaches to semantics, in which meanings were specified in terms of feature lists or strictly hierarchical trees, no formal distinctions were made in semantic memory among concepts, words, properties, features and so on; instead (section 2.1), each entry served both as a concept defined by the system and, through bindings referring to it, as part of the definition of an unlimited number of other entries (*unlimited semiosis*). He felt it especially important that in Model Q, new meanings were created when the system was "nourished by fresh information," and that "further data could be inferred from incomplete data" (pp. 122–125). This was based on the semantic memory processes which create new inferences from the intersections of two or more spheres of activation. Eco's approach to semiotics has been widely accepted (e.g., Malmkjær, 1991). Furthermore, his insight regarding infinite semantic recursion has great relevance to debates regarding the status of features and other decompositional objects still active in semantic memory research (sections 4.6.2, 9.2).

2. The Semantic Priming Effect

In a sentence verification task, Collins and Quillian (1970) noted that RT decreased when a sentence was preceded by another sentence that inferred it. For example, *A canary is a bird* is faster when preceded by *A canary can fly* (which, in Quillian's frame- work, requires the inference A canary is a bird), than when it is preceded by A canary is yellow. Their explanation of this is that the lingering effects of prior activation (of the path from *canary* to *fly* through *bird*) facilitates the later verification of its subpath (ca- nary isa bird); a similar effect was found when the same inference was shared implicitly by two sentences: A canary has wings facilitated A canary has a beak. They also noted that prior reading of sentences containing a given word (i.e., canary) sped RTs for sub- sequent sentences that also contained the same word (but that this effect was based on perceptual identity, not semantic relatedness).

Meyer and Schvaneveldt (1971) presented pairs of letter strings (simultaneously, but most later studies presented a prime string followed by a target string) and subjects were timed as they classified them as "words" or "nonwords" (lexical decision). They found that word decision times were faster when the prime and the target were semantically re- lated, and they considered Collins and Quillian's (1970) concept of activation lingering after spreading through semantic memory as an explanation of the semantic priming effect. A number of studies followed, showing that the effect was robust, but that it could be affected by many experimental factors–for example, Meyer and Schvaneveldt (1976) demonstrated that degraded stimuli, slowing the reading time, increased the semantic priming effect. It became clear that semantic priming is the locus of a variety of mental processes, some dedicated to semantic memory, others to more general cognitive processes; moreover, spreading activation has become the most widely accepted expla- nation (Collins & Loftus, 1975).

1) Controlled versus automatic processes

Neely (1977) found that in addition to the kind of unconscious, automatic spreading activation postulated by Quillian, there was another kind of activation under conscious control that could have a substantial effect on RT and priming effects. This was charac- terized by increased priming (facilitation) from related primes, and negative priming from unrelated primes (inhibition), relative to a neutral prime (usually some kind of “ready” symbol or a nonword). For example, when there was a high proportion of related prime–target pairs and ample time between the presentations of the prime and target, sub- jects adopted an expectancy strategy, which involved constructing a mental list of possible associates to the prime; when the target was in the preparatory set, it could be identified as a word more quickly, and when it was not in the preparatory set, the unsuccessful search lengthened RTs. Neely (1991) proposed an additional strategy called post-lexical semantic matching, which was relevant only for lexical decision; it involved checking for a relationship between the prime and target after the target has appeared; if a relationship was noted, the subject was biased to make a word (as opposed to nonword) response. In general, the overall priming effect was greater when controlled priming processes came into play.

The presence of more than one type of semantic processing may be relevant to Collins and Loftus’s (1975) division of semantic memory into lexical and nonlexical components based on the presence of lexical priming without phonological or orthographic priming, since there is evidence that phonological priming is more dependent on the use of strat- egy and conscious awareness than associative priming (Ober & Shenaut, 1988; Norris, McQueen, & Cutler, 2002) or orthographic priming (Napps & Fowler, 1987). This sug- gests that independent priming from semantic versus phonological primes could be related to differences in strategies used by subjects in performing the task rather than to the existence of separate memory stores (cf. Coltheart, 2004).

2) Methodological implications

Note that Neely’s process-based model represented a substantial shift in emphasis away from earlier models which focused on the structure of concept memory (i.e., fea- tures, semantic networks, spatial models) along with a unitary access process. This shift in emphasis became very important in the neuro-related literature, because a deficit in a controlled process such as expectancy or post-lexical congruency, even though it can produce abnormal semantic priming, is not evidence that the basic structure or contents of the knowledge base is impaired, whereas abnormal semantic priming using methodology that minimizes the use of controlled processing is stronger evidence of a true loss of conceptual knowledge. Neely (1991) identified several methodological variables that affect the degree to which controlled processing is used. The most obvious one is the prime–target stimulus onset asynchrony (SOA): shorter SOAs, below 250 ms or so, do not allow time for expectancy to operate. Continuous priming, that is, a stream of letter strings that are classified or named, reduces post- lexical semantic matching, as does “go/no-go” lexical decision, in which subjects re- spond only to word targets. Masking the prime to below the threshold of identification has been tried as a method to reduce controlled processing, but there is convincing ev- idence that even with masking, subjects can identify the prime often enough to employ controlled strategies. The composition of the stimulus set also affects the use of con- trolled processing. Neely defines the relatedness proportion as the proportion of related word target trials to all word target trials, and the nonword ratio (which applies only to lexical decision) as the proportion of trials with nonword targets relative to all trials other than related prime–target trials. It turns out that when the relatedness pro- portion is high, subjects are more likely to generate expectancy sets, leading to more activation and more inhibition. If the nonword ratio is high, subjects are biased toward making nonword responses when the prime and target are unrelated, resulting in greater inhibition on unrelated trials. Therefore, various combinations of procedural manipulations such as short SOA, fairly low relatedness proportions and/or nonword ratios, and go/no-go responses, can be used to reduce the likelihood that subjects will make use of controlled processing in a semantic priming experiment.

3. Frames, Scripts, Schemata

A critical aspect of Quillian’s framework was that bindings had labels (section 2.1). He distinguished between the proximate bindings in an entry’s plane (analogous to a dictionary entry’s definition), and the more remote bindings reached through spreading activation. Planes had the same form as the structure built up dynamically through spreading activation, intersection, evaluation, and unification; when an entry is acti- vated, activation spreads from the bindings in its plane. Note that the structured, labeled bindings in Quillian’s planes were equivalent to frames (Minsky, 1974), scripts, schemata, and constructions. As an example of the connection to frame theory, Collins and Quillian (1972) adopted the grammatical relations proposed in Fillmore

(1968) for use as bindings identifying the roles of entries acting together in a proposition. Later Fillmore (1976) introduced frame semantics, based on a more highly specified struc- ture related to Quillian's planes (see Petruck, 1996, for a review). Quillian's semantic memory is also connected to cognitive linguistics, an approach to linguistic theory that focuses on the connection between the interaction effect of cognition, the human body, and the environment on language (e.g., Lakoff, 1987; Croft & Cruse, 2004), and to the various construction grammars (e.g., Fillmore, Kay, & O'Connor, 1988).

4. Unification

Quillian implemented the process by which new entries resulting from language un- derstanding became unified on an ad hoc basis in IPL-V and LISP 1.5, the principal artificial intelligence languages available at the time. However, during the same decade, the general unification problem began to receive a great deal of attention (e.g., Robinson, 1965), and new programming languages were developed (most notably, PROLOG) that contained efficient, built-in unification functions. The availability of easy-to-use, efficient unification languages facilitated a number of theoretical advances, often involving data that corresponded more or less to the activated conceptual structures created by Quillian's semantic memory during language understanding. The structures were given labels such as planes, frames, schemata; they contained labeled bindings (similar to Quillian's), and unification operated by locating a binding in two or more structures with the same attribute, whose value was filled in one frame and either filled identically or empty in the others. In this case, the empty values were filled in with the value of the filled binding; when this resulted in all obligatory val- ues being filled, with no contradictions, the entries were unified. As a result of this work, one way that recent models differ from Quillian's is their reliance on more evolved unification functions. See Knight (1989), for a survey of the history, princi- ples, and applications of unification theory.

5. Construction Grammars

Quillian's model was focused on semantic relations among entries, but he was also concerned to some degree with the role of syntax in language use. This was reflected in two ways in his work. In the earlier models, ad hoc form tests were part of the evalua- tion procedure: they rejected intersections that violated syntactic constraints. In later models, he attempted to develop a semantics system that worked in tandem with an in- dependent syntactic network processor (Woods, 1970), such that the two systems inter- acted with each other as they found intersections or constituents. As mentioned above, subsequent natural language-processing systems such as frame semantics and cognitive linguistics combined semantic and syntactic processing in various ways, with a trend away from the syntax/semantics dichotomy. The ultimate development to date of this idea is the construction (Goldberg, 1995; Kay, 2002), which is a complex nexus of bind- ings containing syntactic and semantic information. A construction, in this context, can be a complex sign such as a word (top), a syntactic pattern (NP's NP), or an idiom (X blew X's top). In each case, the structure corresponding to the construction has both syn- tactic and semantic information, and all of the constructions in a system are organized as entries connected via labeled bindings in a semantic network similar to Quillian's se- mantic memory. Perhaps the construction grammar system most relevant to neuro-re- lated applications is embodied construction grammar, which emphasizes the role of the human body, particularly such elements as situation, perception, and the body's config- uration. For example, the fact that we have two hands and ten fingers has influenced syntax and semantics (e.g., Bergen & Chang, 2005). In embodied construction gram- mars, an interaction–activation process works along with unification-like evaluations of structural correspondences. Since the frame-like constructions contain syntactic, semantic, and physical information, these systems can simulate a range of phenomena, including language understanding, inference making, and even perception and manipu- lation of the environment, thereby achieving–surpassing–most of the goals stated but not implemented by Quillian (1969).

6. Concepts and Similarity

Empirical results in the domain of concept knowledge and similarity among concepts have been important influences on the formation of current conceptions of semantic mem- ory. This section briefly considers some of the key phenomena of these overlapping domains.

1) Prototypes and basic levels

As mentioned in section 2.1, Quillian adopted the idea of criteriality from the classic model of Bruner et al. (1956). During the 1970s, several developments superseded the classic model. Probably the most critical work was done by Rosch (1975), who worked within a model similar to the generic "superset hierarchy" version of Quillian's semantic memory. She demonstrated that many concepts appear to be represented in terms of their relation to a possibly abstract prototype, which in turn consisted of a set of features. Items that shared many features with the prototype were more typical exemplars of the concept than items that shared few features. Second, Rosch (e.g., Rosch, Mervis, Gray, Johnson, & Boyes-Braem, 1976) demonstrated that not all entities were accessed with the same degree of naturalness. She showed that there were at least three distinct levels: the basic level (e.g., chair, horse), the most easily accessed, the superordinate level (e.g., furniture, animal), which for Rosch could include any entity with a direct or indirect superset rela- tion to the basic level object, and the subordinate level (e.g., bench, palamino), which could include any entity of which the basic level object has a direct or indirect superset relationship. Note that Rosch's approach introduced an equivocation regarding the terms superordinate and subordinate that has caused some confusion: in most semantic mem- ory models, including Quillian's, there is a constellation of multi-rooted hierarchical bindings such that there can be many levels of logical superordination or subordination connecting two entries; this conflicts with Rosch's three-part division of the entire net- work into basic, superordinate, and subordinate levels. Another possible confusion is that Rosch's prototypes have often been taken to be "best examples" of a category, as opposed to a kind of summary representation consisting of weighted, possibly contradictory fea- tures. For instance, there is no "best example" of a dog that could account for the full range of "dogginess" (various sizes, lengths of hair, ear shapes and sizes, and so on); in- stead, the prototype of dog could be represented in terms of an entry with variably weighted bindings encompassing the entire range of variation, but which would not cor- respond to a particular type of dog. This and other aspects of concept theory are discussed by Murphy (2002).

2) Features, structural alignment

A classic way to represent the mutual similarity among the items in a set, for example, a semantic category, is geometric, such that the distance in space between each pair of items corresponds to the "mental space" (or dissimilarity) between the items. For exam- ple, Rips, Shoben, and Smith

(1973) used multidimensional scaling (MDS) to create two- or three-dimensional plots of items in categories such as bird and animal, and found that MDS distances correlated with category verification RTs. Furthermore, the first two or three dimensions tended to be interpretable as attribute dimensions within the category (e.g., wild versus tame or big versus small). However, Tversky (1977) pointed out that MDS analyses frequently violated basic geometric axioms such as the triangle inequality (e.g., no two-dimensional plot can represent the items bracelet, wristwatch, and clock be-cause bracelet is close to wristwatch but far from clock, while clock is close to wristwatch but far from bracelet). Furthermore, while adding more dimensions resolves this issue, it is extremely rare to be able to interpret more than three MDS dimensions in a meaningful way. Tversky proposed a nongeometric, set-theoretic process model of similarity based on asymmetric, weighted matching of their common and distinctive features. For Tversky, "features may correspond to components such as eyes or mouth; they may represent concrete properties such as size or color; and they may reflect abstract attributes such as quality or complexity" (pp. 15–16). While his feature-matching approach was successful at explaining many phenomena of similarity and conceptual knowledge, there were no ap- propriate bounds on which features were relevant in a certain comparison: in other words, the model was too powerful. An approach originally designed to account for certain phe- nomena of perceptual similarity (Medin, Goldstone, & Gentner, 1993) led to Markman and Gentner's (1993) proposal that the cognitive process used in making a conceptual sim- ilarity judgment was bounded by the attributes of the items being compared, represented in a frame-like structure. In the structured representation approach, which is similar to unification, the structures of the items must first be aligned to the extent possible; com- parisons are made only between corresponding portions of the items. For example, car and truck have many alignable attributes: trucks have two doors while cars have either two or four doors; both have engines, headlights, and steering wheels. However, car and tree have almost no alignable attributes except perhaps extremely general ones like hardness, color, ability to move, and size. It has been found (e.g., Markman & Gentner, 1996) that simi- larity ratings are often more affected by alignable than by nonalignable differences. There is evidence that there can be more than one possible alignment of concepts as a function of past experience, context, and task demands (e.g., Markman, 1999, pp. 289–294).

7. Connectionist Models

The Quillian formulation of spreading activation was an effective information- processing algorithm, but the process of passing a complex marker among entries was not biologically plausible. Newer, neurologically inspired models involved nodes which could be in various levels of activation, connected by links that could either increase (activate) or decrease (inhibit) a destination node as a function of the level of activation of the source node. There were two main families of models based on extensions of this idea: those where the internal nodes corresponded to individual entities (interactive activation models), and those where there was no isolated representation of individual enti- ties (distributed representation models).

1) Interactive activation models

Recall that Collins and Loftus (1975) had proposed a continuous process of spreading activation; to this was added two additional properties, by analogy with neural systems. First, there was a process of inhibition (negative activation), such that the activation of one entry caused the activation of another one to decrease. Second, similar entries could be organized into mutual inhibitory sets (lateral inhibition). When entries were con- nected with bindings that could either increase or decrease the activation of other entries with various strengths, it was found that activating certain entries externally caused the system to enter a state of disequilibrium which lasted for a time, but eventually resolved into a stable state corresponding to the result of a Quillian spreading-activation plus evaluation cycle. In particular, lateral inhibition caused all but one of a set of mutually interconnected similar entries to be suppressed, resulting in clear discrimination between activated and not activated states. "Interaction–activation" models were used initially to perform such tasks as letter recognition (McClelland & Rumelhart, 1981) and word–sense disambiguation (Cottrell & Small, 1984), and have been extended to many other domains.

2) Distributed representation models

Quillian's semantic memory was the first computer-based connectionist model applied to a wide body of problems. Yet there was some concern that because the individual entries and bindings were set up by the experimenter, an element of bias could be intro- duced. In response, a new class of connectionist models were developed whose inputs, outputs, and architecture were specified by the experimenter, but whose internal structure was a tabula rasa. These systems tended to resemble interactivation–activation models, with units organized into massively interconnected planes, and with feed-forward con- nections from one plane to the next. Some of the these models had a feedback mecha- nism resembling a clocked state machine, where some of the outputs of cyclen were gated back to become inputs in cyclen+1 (recurrence). The systems were programmed using a mechanism known as back-propagation: the experimenter determines for each possible input what outputs are correct, and cycles through a presentation of each input (or input sequence in the case of recurrent architectures); the degree to which the outputs differ from the correct output is propagated back along its inputs, at each point, the weights of the connections are changed slightly. This process is repeated until the outputs are close enough, as defined by the experimenter, to the correct outputs for each input configura- tion. Because of the iterative back-propagation programming sequence, the distinctions are continuous (graded) rather than discrete; this also mimics human performance. While the input and output units are programmed and read explicitly, the intermediate ("hid- den") units change their settings implicitly. These systems have been demonstrated to be extremely powerful, and furthermore to exhibit a critical element of human cognition, graceful degradation. See Rogers and McClelland (2004) for an extensive overview of distributed representation models.

Because Quillian-style models of semantic memory have symbolic entries correspon- ding to concepts, lexical items, and similar mental objects, they are known as localist connectionist models in contrast to distributed representation models, in which each in- terconnected unit is involved to some degree with the system's response to each input. There has been considerable debate regarding the relative value of localist and distributed representation models as vehicles for semantic analysis. Three issues are particularly im- portant in this regard. First is the issue of systems that must be programmed versus sys- tems that learn. While it is true that learning is central to back-propagation distributed representation systems, localist systems have been developed with the ability to allocate new entries and thereby to acquire new concepts (see Page, 2000, and commentaries for a thorough overview of this issue; see also Waskan, 2001). Second, some have held that distributed representation systems are more like actual brain function, in fact, this is why distributed representation systems are also known as “neural networks.” However, the connection tends to be one of inspiration more than a realistic simulation of the brain–for example, no biological system uses anything like back-propagation. Third, since distrib- uted representation systems learn only to produce outputs corresponding to inputs, it is difficult for them to bind together arbitrarily complex, nested propositional structures (the “binding problem”; see Roskies, 1999, and articles in the same volume). Finally, as Page (2000) points out, many purportedly distributed models have important localist structure such as in the arrays of input and output units; that is, they are actually localist- distributed hybrids.

Distributed representation models incorporate assumptions about how degradation of semantic memory could occur in brain-damaged populations, primarily by analogy with graceful degradation. McClelland (1987, p. 472) uses three methods to analyze how a distributed representation system is degraded due to damage: by randomly deleting input nodes, by randomly destroying connections from a unit, and by adding random noise to connection weights. Due to the fact that every unit participates in every input–output mapping to some degree, most of the programming is preserved even with fairly exten- sive damage. This has led to a large body of work in which various sets of knowledge are programmed (via back-propagation) into a distributed representation system, followed by network damage thought to resemble brain damage due to pathology; the resulting errors in input–output mapping are then compared to the performance of clinical subjects.

Some cognitive neuroscientists, including Tyler and Devlin (e.g., Devlin, Gonnerman, Andersen, & Seidenberg, 1998; Tyler, Moss, Durrant-Peatfield, & Levy, 2000; Randall, Moss, Rodd, Greer, & Tyler, 2004) have developed distributed representation models of semantic/conceptual knowledge that make assumptions not only about the differences in feature representation for living things versus nonliving things (as per Farah & McClelland, 1991; Tippett & Farah, 1994), but also about the importance of distinctive- ness of features and of the correlations among features. A selective review of this work is in section 5.4.

D. Semantic Memory and the Brain

The remainder of this chapter will focus on neuro-related aspects and issues of (generic) semantic memory. In this section, we critically review five major models of the brain underpinnings of semantic memory. These models are based on data from case studies and group studies of neuropsychological patients, including those with category specific impairment (CSI; see Table 1), AD (see Table 2), and semantic dementia (SD; see Table 3). Neuropsychological and structural neuroimaging findings with these patient groups, and more recently, functional imaging findings with normal controls, have provided the empirical database upon which these models have been constructed. Not surprisingly, more recently developed models have greater scope (in terms of phenomena to be explained) and specificity (in their predictions) than earlier models. Although some of these models have proposed multiple systems/subsystems of semantic memory, they all assume that semantic knowledge is organized on the basis of representational con- straints imposed by the brain.

Table 1

Overview of category-specific impairment (CSI) and semantic memory.

Description. Patients with category-specific impairment (CSI) have a disproportionate impair- ment for particular semantic categories or domains (*e.g., living versus nonliving*). A landmark publication on CSI was authored by Warrington and Shallice (1984) who described four herpes encephalitis patients (young and middle-aged adults), with bilateral temporal damage, who had global amnesia, were generally impaired on picture naming and word definition tasks, but were very disproportionately impaired when the stimulus items represented animals or plants (i.e., liv- ing things) as compared to artifacts (*e.g., tools, musical instruments*). The CSI cases reported since 1984 have included some with an artifacts deficit, although a *living-things* deficit is much more common. There have also been occasional reports of patients with selective deficits for spe- cific categories such as *fruits, vegetables, or animals* (e.g., Caramazza & Shelton, 1998; Hart & Gordon, 1992).

Neuropsychological considerations & confirmation of a semantic deficit. There are two methodo- logical issues that are critically important in making the determination that a patient has CSI. First, it must be demonstrated that the performance deficits on the semantic tasks are not due to percep- tual problems (e.g., deficits in perceptual-level processing of pictures or words) or word retrieval deficits; otherwise, it will be unclear as to whether the performance deficits are truly semantic/con- ceptual in nature, as opposed to being restricted to a particular input or output route to/from seman- tic memory. (This is also important in the evaluation of semantic/conceptual knowledge in AD or SD.) Second, the stimulus materials used from the various categories/domains being tested must be matched on variables that can affect performance (e.g., frequency of occurrence for word stim- uli, visual complexity for picture stimuli). Otherwise, it could be the case that the category or do- main for which performance is impaired happens to have stimuli that are more difficult to encode, recognize, or name.

Neurological correlates. Saffran and Schwartz (1994) describe the patients with disproportion- ate impairment of *living things* as most commonly having bilateral temporal lobe damage due to herpes encephalitis, and, less commonly, having temporal lobe (and sometimes, additionally, frontal lobe) damage due to infarct(s) or a degenerative disorder of unknown origin. In contrast, Saffran and Schwartz describe the patients with disproportionate impairment *of nonliving things* as most commonly having frontoparietal lesions of cerebrovascular origin. A recent review by Capitani et al. (2003) encompasses 61 CSI patients with deficits for one or more biological categories, and 18 CSI patients with deficits for one or more categories of *nonliving things*; the lesion sites were generally in temporal cortex for the former subgroup, and in frontoparietal or (less commonly) temporal cortex for the latter subgroup.

Table.2

Overview of Alzheimer’s disease (AD) and semantic memory.

Description. Alzheimer’s disease (AD) is the most common type of dementia, accounting for about one-half of all dementia cases. “Probable AD” and “possible AD” are diagnosed in vivo by uni- versally accepted exclusionary criteria (McKhann et al., 1984). A definitive diagnosis of AD can only be made on the basis of autopsy findings, when brain tissue samples show the requisite con- centration of senile plaques and neurofibrillary tangles; when the criteria for “probable AD” are followed, the autopsy data confirm this diagnosis in about 95% of cases (Morris, McKeel, Fulling, Torack, & Berg, 1988; Tierney et al., 1988).

Neuropsychological profile. Episodic memory impairment is the hallmark feature of AD; memory for recent events is impaired even in the earliest stages of the disease. Anomia is also evident, as are difficulties in attentional functioning, in the early stages. As the disease progresses, widespread cognitive impairments occur.

Semantic memory impairment. The nature and extent of deficits in semantic memory associated with mild-to-moderate AD have been the subject of much debate (see Ober, 1999, and the seven thematic articles that follow, and which comprise this Journal of the International Neuropsychological Society symposium, for a broad-ranging discussion of the issues). AD patients are impaired in their performance on standard neuropsychological tests of semantic memory, such as confrontation naming and verbal fluency (e.g., Martin & Fedio, 1983; Ober, Dronkers, Koss, Delis, & Friedland, 1986; Thompson-Schill, Gabrieli, & Fleischman, 1999). However, there are laboratory tests of semantic memory that show normal performance in AD (these findings are reviewed in this chapter).

Neurological correlates. The structural and functional brain abnormalities associated with the earlier stages of AD are overwhelmingly in posterior (temporal–parietal) as opposed to anterior neocortex, and are particularly evident in the hippocampus and the medial temporal lobe (e.g., Braak & Braak, 1991; Jack et al., 1997; Parks, Haxby, & Grady, 1993). Only as the disease pro- gresses into the moderate and then severe stages, will significant atrophy and decreased function occur in more widespread areas of the neocortex, including frontal, anterior temporal, and lat- eral temporal areas (e.g., Parks et al., 1993; Scahill, Schott, Stevens, Rossor, & Fox, 2002). The contrasting brain pathology for AD compared to SD in the early stages of the disease, is consis- tent with episodic memory being affected disproportionately to semantic memory in AD, whereas the reverse holds for SD.

1. Sensory–Functional Theory

Sensory-functional theory was formulated by Warrington and colleagues (Warrington & Shallice, 1984; Warrington & McCarthy, 1987), who observed that the pattern of intact and impaired categories in CSI patients sometimes, but not always, conformed to a living things versus nonliving things dissociation. For example, some CSI patients showed impairment for musical instruments as well as for living things. The assumptions of sensory-functional theory are: (1) semantic memory is organized into modality-specific subsystems (e.g., visual/perceptual, functional/associative); and (2) the ability to recognize (and name) living things is relatively more dependent on.

Table.3

Overview of Semantic dementia (SD) and semantic memory.

Description. In a landmark study, Warrington (1975) documented a pattern of neuropsychologi- cal test performance, for three patients with degenerative brain disease which led her to conclude that these patients were suffering from semantic memory deficits. These patients had significant problems with word retrieval and word comprehension, and exhibited impoverished knowledge of many semantic domains. Other aspects of language and cognition, however, including day-to- day event memory were relatively well preserved. Warrington drew upon Tulving's (1972) distinction between episodic and semantic memory, in describing the deficits of these patients as semantic memory deficits. Similar cases were reported in subsequent years; and Snowden, Goulding, and Neary (1989) first coined the term semantic dementia for these patients. SD is rare, in comparison to most other types of dementia.

Neuropsychological profile. John Hodges and his colleagues have published extensively on SD. Their research has included thorough neuropsychological testing of language, memory, and visuo-spatial functioning (e.g., Hodges, Patterson, Oxbury, & Funnell, 1992; Hodges, Patterson, & Tyler, 1994). SD presents with a language disorder that worsens gradually over time, and is usually not accompanied by other cognitive or behavioral deficits for two or more years from onset in contrast to other, more typical dementias (Mesulam, 2001). It has, in fact, been consid- ered by many clinicians to be a fluent form of primary progressive aphasia (see Grossman & Ash, 2004, for a review of fluent versus nonfluent primary progressive aphasia, with the former being equated to SD).

Semantic memory impairment. Detailed assessments of semantic knowledge, using picture as well as word stimulus materials, and assessing visually based semantic knowledge (e.g., Hodges et al., 1994, 1992) support the idea that SD is not just a language disorder, but an actual disorder of semantic memory.

Neurological correlates. SD has been termed a temporal variant of frontotemporal dementia, in that SD patients have extensive atrophy (greater on the left than right) in the polar, lateral, and inferior regions of the temporal lobe; this is in contrast to the marked atrophy in frontal regions, without specific semantic memory impairment, seen in the frontal variant of frontotemporal dementia (Galton et al., 2001).

Visual information, whereas the ability to recognize (and name) nonliving things is rel- atively more dependent on functional information. These assumptions lead to three pre- dictions: (1) dissociations will not occur within the domain of living things, since the same semantic subsystem (visual) is critical for all living things; (2) patients with cat- egory-specific deficits will also have deficits for the modality/type of information which is critically involved in recognizing items from the impaired category (e.g., liv- ing things deficits should be accompanied by visual–perceptual knowledge deficits across all categories/domains); and (3) patients with disproportionate deficits for a given modality/type of knowledge will also have a disproportionate deficit for the cat- egory/domain that depends on that type of knowledge. None of these predictions, how- ever, have been fulfilled (for reviews of the relevant CSI data see Capitani, Laiacona, Mahon, & Caramazza, 2003; Caramazza &

Mahon, 2003; and Caramazza & Shelton, 1998). The original formulation of sensory-functional theory has been largely aban- doned. However, revised formulations of sensory-functional types of theories have been developed; one of these is Damasio's (1989) convergence zone theory, upon which Simmons and Barsalou's (2003) conceptual topography theory is partly based (section 5.5).

2. Sensory–Motor Theory

The sensory–motor theory of Martin, Ungerleider, and Haxby (2000) proposes that conceptual knowledge is represented in the brain according to the features that define the object concepts (e.g., tools, animals) under study. Moreover, this theory assumes that semantic memory is functionally unitary and distributed over modality-specific repre- sentations. This is in contrast to the assumption of sensory-functional theory that conceptual knowledge and modality-specific representations are functionally and neuro- logically dissociable, and that there are subsystems of semantic memory. Sensory–motor theory is based, in large part, on neuroimaging data with normal subjects in which: (1) retrieval of the color or action associated with given objects activated inferior and supe- rior, respectively, regions of the temporal lobe, areas known to mediate color versus action/motion perception (Martin, Haxby, Lalonde, Wiggs, & Ungerleider, 1995); and (2) naming of tools (relative to animals) activated an area of left temporal lobe overlapping with that activated by action naming in the Martin et al. (1995) study as well as a region of the left premotor cortex known to be involved in imaging of actions, whereas naming of animals (relative to tools) activated inner regions of occipital cortex, bilaterally (Martin, Wiggs, Ungerleider, & Haxby, 1996). Additional support for models in which partial representations of a given object are purportedly stored in or near the primary sen- sory and motor areas that are involved in perception and learning of the object's features comes from numerous functional imaging studies in which retrieving specific types of object attributes (e.g., color, action, and visual form) activated the same brain areas that have been shown to mediate perception of those attributes (e.g., Chao & Martin, 1999; Howard et al., 1998; Kable, Lease-Spellmeyer, & Chatterjee, 2002; Oliver & Thompson- Schill, 2003). Moreover, Martin et al. (2000) and other cognitive neuroscientists (e.g., Badre & Wagner, 2002) have proposed that specific regions of the left prefrontal cortex and the anterior temporal cortex have particular roles in retrieving, maintaining, and selecting semantic information (i.e., in “working with semantic representations”). For a review of the specific brain regions that are engaged in access, selection, and retrieval of semantic representations, see Thompson-Schill, Kan, and Oliver (2006).

3. The Domain-Specific Hypothesis

The domain-specific hypothesis, a multiple systems/subsystems model, holds that knowledge of a category is not distributed among the sensory–motor systems involved in processing category exemplars, but in category-specific brain systems that are “down- stream” from sensory–motor processing (Caramazza & Shelton, 1998). This type of organization is said to be due to evolutionary pressures for rapid and efficient processing of selected semantic domains, such as animals, plants, conspecifics (members of the same species), and possibly, artifacts such as tools. The domain-specific hypothesis (as delineated in Caramazza & Mahon, 2003) predicts that (1) conceptual deficits should typ- ically affect just one of the (evolutionarily significant) categories/domains, and if the sys- tem for one domain is damaged it will not be possible for the function of this system to show “recovery” based on the functioning of another such system; (2) there is no neces- sary association between a deficit for a type/modality of knowledge and a conceptual deficit for a specific category/domain; and (3) perceptual (i.e., pre-conceptual/semantic) stages of object recognition may be functionally organized via domain/category con- straints, as are conceptual/semantic stages of object recognition. As per Caramazza and Mahon, evidence in support of the first prediction can be found in Farah and Rabinowitz (2003); evidence in support of the second prediction comes from living things as well as nonliving things deficit cases showing equivalent impairments of visual/perceptual and associative/functional knowledge (e.g., Caramazza & Shelton, 1998; Laiacona & Capitani, 2001; Moss & Tyler, 2000; Samson, Pillon, & Wilde, 1998); evidence in support of the third prediction comes from cases showing equivalent deficits in visual/perceptual and functional/associative knowledge of living things, in the face of visual agnosia for living things (e.g., Caramazza & Shelton, 1998; Laiacona, Barbarotto, & Capitani, 1993). Caramazza and colleagues state that there is much indirect evidence for a very limited range of categories (as described above) being affected by domain-spe- cific deficits;

this evidence is said to be the pattern of category deficits across patients, as reviewed in Capitani et al. (2003).

4. Conceptual Structure Theory

The conceptual structure theory (CST) of Tyler and colleagues (Tyler & Moss, 2001; Randall et al., 2004) is a distributed, connectionist model of semantic knowledge. CST assumes that category-specific semantic deficits (i.e., in CSI patients) as well as the non- specific deficits seen in other neuropsychological populations (including AD and SD) are the result of random damage to a conceptual/semantic system which is not organized a priori by object domain or feature modality. CST incorporates the following specific assumptions: (1) living things have more shared features (and thus, fewer distinctive features) than nonliving things; (2) for living things, biological function information is highly correlated with shared perceptual properties (e.g., can see – has eyes); (3) for arti- facts, function information is highly correlated with distinctive perceptual properties (e.g., cuts as the function of knife, via its blade); (4) semantic categories, within (living versus nonliving) domains, differ in their structure (e.g., vehicles are less typical than other nonliving categories in having more properties overall and more shared versus distinctive properties, than do tools); and (5) features that are highly correlated with other features will be more resistant to damage than features that are not highly correlated. Assumptions 1–4 have received support from property generation norming, property verification, and other experimental work with large groups of young normals (YN). Assumption 5 has been evaluated with several semantic-deficit case studies and with computational models, with mixed support. (For reviews of this work see Randall et al., 2004; Tyler et al., 2000; and Tyler & Moss, 2001.) There are other similarity and correlation-based distributed models of conceptual knowledge that have many of the same assumptions as CST (e.g., Devlin et al., 1998; McRae, de Sa, & Seidenberg, 1997); CST is the only such model, however, that incorporates assumptions about the interaction between distinctive- ness and correlation within living and nonliving domains. A key prediction of CST, stem- ming from the assumptions that living things have relatively more correlated features and highly correlated features are more resistant to disruption, is that a disproportionate deficit for living things will be observed when the damage to the semantic system is mild, whereas a disproportionate deficit for nonliving things will arise only if the damage is se- vere enough that all that is left are some of the highly correlated, shared features of liv- ing things. In contrast, the correlated-structure model of Devlin et al. (1998) assumes that disrupting access to a given feature will disrupt access to highly correlated features, thus predicting that a disproportionate deficit for living things will occur at severe levels of damage. The evidence on this issue has been obtained mainly from patients with AD, and is quite mixed/inconclusive (e.g., Gonnerman, Andersen, Devlin, Kempler, & Seidenberg, 1997; Garrard et al., 2001; Zannino, Perri, Carlesimo, Pasqualetti, & Caltagirone, 2002). One limitation of CST is that it does not account for patients with disproportionate deficits for nonliving things in the face of relatively intact performance for living things (i.e., in patients who are in the earlier versus later stages of disease). Although this pat- tern of deficits is far less common than the pattern involving an early living things deficit, it does occur (e.g., patient “JJ,” described by Hillis & Caramazza, 1991); the domain- specific hypothesis can, of course, account for nonliving things deficits, even in mild-to- moderately (as opposed to severely) impaired patients.

A series of neuroimaging studies by Tyler and colleagues was designed to reveal pat- terns of brain activity specific to category and/or domain (living versus nonliving) in nor- mal subjects, when stimulus and task characteristics were carefully controlled across these domains; such patterns would support the domain-specific hypothesis as opposed to the CST, whereas the absence of such patterns would support CST (or at least a uni- tary, distributed model of semantic memory). These papers used both fMRI and PET, lex- ical-decision as well as category-judgment tasks, picture as well as word stimuli, verb as well as noun word stimuli, and several living (e.g., animals) versus nonliving (e.g., tools) categories (Devlin et al., 2002; Pilgrim, Fadili, Fletcher, & Tyler, 2002; Tyler et al., 2003). Findings across these studies showed robust activations for the semantic/catego- rization tasks as compared with baseline tasks; these activations were mainly in left frontal (particularly inferior frontal) and left temporal regions. However, in none of these studies were there differential activations for the processing of the different categories (living versus nonliving; different categories within the living versus nonliving domains). Tyler and colleagues interpret their overall findings as supporting a model in which conceptual knowledge is represented within a unitary, distributed semantic/conceptual system, and as inconsistent with either the domain–specific hypothesis or with the sensory-motor theory hypothesis that different neural networks are responsible for pro- cessing different types of information that are associated with different categories of knowledge (e.g., tools versus animals). It is important to note that Mahon and Caramazza (2003) take issue with the argument that functional neuroimaging results indicating category-differential (as well as category-identical) patterns of activation, as opposed to those yielding category-selective patterns of activations, are not interpretable within a domain-specific framework (as argued by Moss & Tyler, 2003; Tyler & Moss, 2001). Mahon and Caramazza make the opposing argument that functionally discrete processes do not have to be carried out by non-overlapping neural regions; rather, they can be car- ried out by overlapping regions (e.g., Martin & Weisberg, 2003). A potentially influen- tial distributed representation model of semantic deficits is that of Rogers et al. (2004). This interdisciplinary model, which is focused on SD, has an emphasis on: (1) the man- ner in which semantic representations emerge from the interactions among modality-spe- cific representations of objects, and (2) accounting for multiple aspects of normal versus impaired performance on tests of semantic/conceptual knowledge.

5. Conceptual Topography Theory

Simmons and Barsalou (2003) have developed a theory of conceptual knowledge, con- ceptual topography theory (CTT), which integrates aspects of both neural-structure and correlated-structure theories. Simmons and Barsalou propose that sensory-functional in- formation, evolutionarily significant categories, and statistical relationships between cat- egories and their properties are all important aspects of the organization of conceptual knowledge. They utilize convergence zone theory (Damasio, 1989; see also Damasio & Damasio, 1994) as a basis for synthesizing aspects of sensory-functional, domain-spe- cific, and conceptual structure theory. Convergence zone theory begins with the widely accepted assumption that when an object is perceived, it activates feature detectors in relevant sensory–motor areas (in the case of vision, these could be, e.g., for shape, ori- entation, or color); these systems of detectors (called feature maps) are organized hierar- chically for vision as well as for other modalities. The key innovation of convergence zone theory is its explanation of how the states of activation within feature maps are stored. Damasio proposes that the neurons in a nearby association area (conjunctive neu- rons) bind the pattern of activated features for use later.

These association areas are referred to as convergence zones and they are proposed to be organized in multiple hier- archical levels, with the convergence zones that are located near specific sensory–motor areas capturing patterns of activation relevant to that modality (e.g., visual, auditory, motor areas), and the convergence zones that are located away from specific sensory–motor areas are involved in capturing increasingly higher levels of pattern acti- vation, including convergence zones that integrate information across the highest levels of modality-specific convergence zones. Once the feature maps have been established in the convergence zones, the conjunctive neurons in these zones can re-enact the patterns of activation for instances of given concepts without bottom-up sensory stimulation, that is, via recollection/imagery (note the similarity to Quillian-style activation and subse- quent unification of subnetworks). Simmons and Barsalou (2003) explain how the addi- tion of two principles to Damasio's (1989) convergence zone theory can result in CTT, which can explain what is known about conceptual deficits, while synthesizing the three dominant theories (sensory-functional, domain-specific, and conceptual structure). It should be noted that the Simmons and Barsalou article is an extension of Barsalou (1999) in which convergence zone theory was applied to Barsalou's perceptual systems theory of normal conceptual/semantic knowledge; the Simmons and Barsalou extension is meant to explain conceptual/semantic deficits in neuropsychological populations. The two new principles are (1) the similarity-in-topography (SIT) principle, and (2) the vari- able dispersion principle (which is a corollary of the SIT). Per Simmons and Barsalou (2003, p. 457) "The SIT principle concerns the organization of the conjunctive neurons in CZs (convergence zones). Essentially the SIT principle claims that categorical struc- ture in the world becomes instantiated in the topography of the brain's association areas. Specifically, the SIT principle states that: The spatial proximity of two neurons in a CZ reflects the similarity of the features they conjoin. As two sets of conjoined features be- come more similar, the conjunctive neurons that link them lie closer together in the CZ's spatial topography." The variable dispersion principle assumes that conjunctive neurons for a category are dispersed in clumps, with a given clump containing conjunctive neu- rons that are utilized for more than one category. In other words, there is low dispersion for categories with instances that have high similarity (e.g., mammals) and high disper- sion for categories with instances that have relatively low similarity (e.g., artifacts). This relates to category-specific semantic deficits, because a lesion in an area where the clus- ters for a category are tightly localized will be more likely to lead to disruptions in per- formance for that category, than when the lesion occurs in an area where there is not this type of localization (i.e., where there is a cluster that is one of many widely distributed clusters for a category/domain). Simmons and Barsalou argue that CTT is quite success- ful at modeling a wide range of conceptual deficits in neuropsychological populations, as well as accounting for a wide variety of phenomena in conceptual/semantic processing in normal subjects. They do note, however, that additional direct evidence for the SIT principle, that is, data concerning the relation between conceptual similarity and topog- raphy within convergence zones, is required. This type of data will most likely have to come from the next generation of high-resolution, event-based neuroimaging studies of conceptual processing, with normal as well as neuropsychological populations. Neither critiques of, or data contradictory to, the CTT have yet appeared in the literature.

An intriguing source of empirical support for Simmons and Barsalou's (2003) SIT prin- ciple is based on the observation that certain "mirror neuron" cells known to be involved in performing certain actions become activated when the subject observes others perform them. For example, the performance of goal-directed actions by humans as well as mon- keys activates a network including premotor, motor, and posterior parietal regions; the ob- servation of another individual performing those actions activates the same premotor and posterior parietal regions, but not the motor regions (e.g., Iacoboni et al., 1999). Another example is that when humans observe videos of others inhaling odorants that produce feel- ings of disgust, the same regions of the anterior insula and anterior cingulate are differen- tially activated (compared to neutral odorants), as when the disgust-producing odorants are actually inhaled (e.g., Wicker et al., 2003). A final example is the recent (human) find- ing, that a mirror neuron system in the inferior frontal cortex differentially responds to the observation of grasping actions in an appropriate, meaningful context from the observa- tion of such actions in the absence of context (Iacoboni et al., 2005). All of these findings converge on the notion of highly overlapping brain regions being involved in the percep- tual–motor as well as conceptual/semantic aspects of knowledge.

Semantic memory research was for many years dominated by cognitive psychologists who generally were not concerned with neural organization. In cognitive neuropsychology, there is a history of studies investigating patients with semantic deficits (Warrington & Shallice, 1984). However, for a number of years, this line of research was divorced from semantic memory research using normal adult participants. With the advent of neuroimaging techniques, fMRI in particular, research on the neural organization of semantic memory blossomed. Researchers have long known that brain regions responsible for perception tend to be specialized for specific sensory modalities. Given that perception is distributed across specialized neural regions, one possibility is that conceptual representations are organized in a similar fashion. For the past 40 years, Paivio (1971) has advocated a form of modality-specific representations in his dual-coding theory. Furthermore, studies of patients with category-specific semantic deficits have been used as a basis for arguing for multimodal representations for the past 25 years or so. In early work, Warrington and McCarthy (1987) put forward their sensory/functional theory to account for patterns of category specific impairments of knowledge in patients with focal brain damage. The basic assumption is that living things depend primarily on visual knowledge, whereas although visual knowledge is also important for nonliving things, knowledge of an object's function is primary. Building on Allport (1985), recent research has used analyses of large scale feature production norms to extend the sensory-functional theory to other senses and types of knowledge, and move beyond the binary living-nonliving distinction (Cree & McRae, 2003). There do remain some accounts of category-specific semantic deficits that are amodal (Caramazza & Shelton, 1998, Tyler & Moss, 2001), but even these researchers

Semantic Memory 6 have begun to find support for theories in which knowledge is tied to modality-specific brain areas (Mahon & Caramazza, 2003; Raposo, Moss, Stamatakis, & Tyler, 2009). The behavioral and neuropsychological evidence in favor of grounded semantics is corroborated by recent neuroimaging studies supporting a distributed multimodal system. A few researchers have used evoked response potentials to investigate this issue (Sitnikova, West, Kuperberg, & Holcomb, 2006), but the vast majority of studies have used fMRI. For example, Goldberg, Perfetti, and Schneider (2006) tied together previously reported neuroimaging evidence supporting modally bound tactile, colour, auditory, and gustatory representations. They found that sensory brain areas for each modality are recruited during a feature verification task using linguistic stimuli (e.g., banana-yellow). The same pattern emerges in single word processing. Hauk, Johnsrude, and Pulvermüller (2004) showed that reading action words correlates with activation in somatotopically corresponding areas of the motor cortex (lick activates tongue regions while kick activates foot regions), indicating that word meaning is modally distributed across brain regions. Furthermore, within brain regions that encode modality-specific, possibly feature-based representations, some studies suggest a category-based organization (Chao, Haxby, & Martin, 1999). Finally, some studies have shown that semantic representations are located just anterior to primary perceptual or motor areas, whereas others have found evidence for activation of primary areas (see Thompson-Schill, 2003). In summary, there is a large amount of converging evidence supporting a distributed multimodal semantic system (for thorough reviews, see Binder, 2009; Martin, 2007). Perhaps one the most important remaining issues concerns the fact that people's concepts are not experienced as a jumble of features, disjointed across space and time, but instead are experienced as coherent unified wholes. Multimodal feature-based theories therefore need to

Semantic Memory 7 include a solution to the binding problem, specifying how representational elements are integrated into conceptual wholes, both within and between modalities. One solution involves temporal synchrony of neuronal firing rates (von der Malsburg, 1999). Semantic representations may be integrated by coincidental firing rates of distributed neural populations. However, the most frequently invoked solution is based on the idea of a convergence zone, which can be considered as a set of processing units that encode coincidental activity among multiple input units (Damasio, 1989). In connectionist models, a convergence zone may be thought of as a hidden layer (Rogers et al., 2004). Because they encode time-locked activation patterns, an important property of convergence zones is that they transform their input, rather than just repeat signals. In this way, successive convergence zones build more complex or abstract representations. Current theories of multimodal semantic representations incorporate either single convergence zones, as in

Patterson, Nestor, and Rogers' (2007) anterior temporal lobe hub theory, or a hierarchy of convergence zones encoding information over successively more complex configurations of modalities (Simmons & Barsalou, 2003). At the moment, it is unclear which of these hypotheses is correct. In summary, recent research supports the idea that semantic representations are grounded across modality-specific brain regions. Researchers are working toward fleshing out details of precisely what these regions encode, the degree to which sub-regions are specific to types of concepts, and how semantic representations are experienced as unified wholes. Furthermore, the vast majority of research has been conducted on concrete concepts, so research on other concepts, such as verbs or abstract concepts, will play a key role over the next few years.

E. Semantic network theory

Semantic networks are collections of nodes linked together by labeled relational links. Each node typically represents a sin-gle concept, and hence these models are referred to as having localist representa-tion schemes. The meaning of a concept is represented through a set of pointers to other nodes. A goal of this type of mod-eling is to determine how to link up the nodes such that the resultant knowledge structures can be used to produce realistic semantic inferences. The implementations that have best stood the test of time are those of Ross Quillian (e.g., Quillian, 1962; 1967; 1968; 1969), who was concerned with understanding both natural language and memory. Quillian's early models worked by instantiating specially coded dictionary definitions into networks of nodes and examining how inferences could be drawn from the intersections of paths emanating from target nodes.

Collins and Quillian (1969) realized that if they included assumptions about effi-ciency of storage and the length of time it should take to move between nodes, it would be possible to derive predictions about how humans retrieve information. They suggested that conceptual information was stored in a hierarchy, with more general concepts (e.g., animal) at the top, and more specific concepts (e.g., canary) at the bot-tom (see Figure 13.1). Concepts were defined in two ways: as a set of features held within each concept node, and in the set of pointers to other nodes. Properties of concepts were stored at the highest node in the hierarchy for which the property held true for all concepts below (e.g., <has wings> was stored at the bird node, but not the animal or canary node), thus implementing a form of cogni-tive economy.

Predictions regarding the length of time it should take participants to verify state-ments could be generated directly from the model. These were tested most thoroughly through use of the sentence verification task in which participants were asked to verify the truth/falsity of simple sentences (e.g., “A canary is a bird.”). This task could be simulated in the Collins and Quillian framework by starting at a node and search-ing properties in nodes and along relational links to other nodes until the information necessary to evaluate the statement had been found. Collins and Quillian found, as predicted, that it took longer for people to verify statements that required longer searches (e.g., traveling two nodes) than statements that required shorter searches (e.g., traveling one node).

The Collins and Quillian framework also provided a mechanism through which infor-mation could be inherited. If the system needed to learn about a new concept, then a node for that concept could be attached at the appropriate level of the hierarchy, and the concept would automatically inherit all of the appropriate information about mem-bers of that category that were stored at higher nodes.

Initial behavioral evidence appeared to support both the ideas of hierarchical orga-nization and inheritance through cognitive economy (e.g., Collins and Quillian, 1969), but hierarchical network theory did not hold up well to further investigation. A series of findings convincingly demonstrated that the strength of relation between a concept and property, or concept and concept, was more important in determining verification latency than was distance in the hierar-chy (e.g., Conrad, 1972). Furthermore, the model could not explain typicality effects, such as why people are faster to verify that a robin is a bird than that a chicken is a

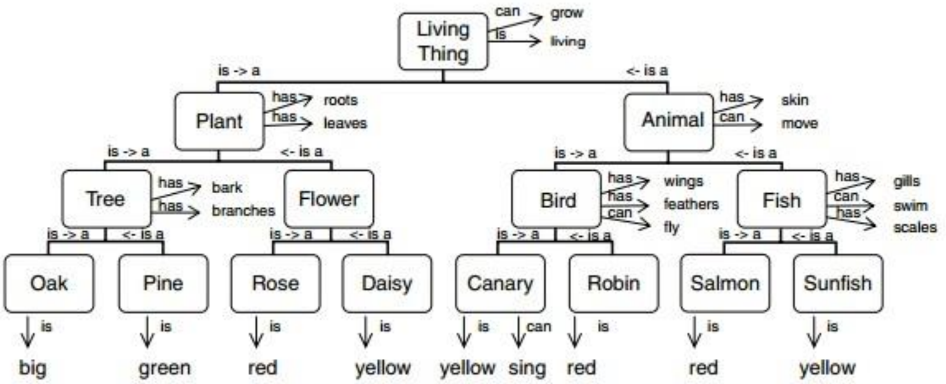


Figure 13.1. Architecture of the hierarchical network theory as proposed by Collins and Quillian (1969).

bird (Rosch and Mervis, 1975). Additionally, it was not clear how one would decide where in the taxonomy to store concepts that belonged to more than one category (e.g., knife), or when to add a new node for items that were similar, but not identical, to those already represented. These obser-vations, along with others, were used to argue against semantic memory as a strict taxonomic hierarchy.

Spreading activation theory (Collins and Loftus, 1975) was proposed as an alterna-tive and was framed as a semantic network without hierarchical organization. It was used to account for a number of behavioral phenomena that posed problems for hierar-chical network theory. But this power came at a cost, as it was ultimately determined that the model was too flexible and could be used to account for just about any pat-tern of data (Johnson-Laird, Herrmann, and Chaffin, 1984). This flexibility came mainly from the fact that there were no constraints on which nodes could be connected to which other nodes, and more importantly, no manner for determining the strengths of weights between nodes.

Semantic network theory was the first major computational approach designed to investigate semantic memory. It succeeded in inspiring over a decade of behavioral research and in promoting the need for future mod-els to account for inheritance and typicality effects. The modeling framework itself is quite simple and provides a language with which one can easily discuss predictions. Yet despite this, the framework is unsatisfying, because the concessions needed to allow it to account for the known behavioral data leave the approach too unconstrained, and hence unviable as a research framework within which to study human cognition. A small number of theories were proposed as alternative explanations of the sentence verification data (e.g., McCloskey and Glucksberg, 1979; Smith, Shoben, and Rips, 1974), but none were proposed as grand the-ories of semantic memory on the scale of hierarchical network theory or spreading activation theory. Despite the limitations, semantic network models are still under development, although more as a means to implement an efficient knowledge system than for generating novel predictions. The various instantiations of ACT, for example, include a spreading activation-like seman-tic component (Anderson, 1983). A second example, although not exactly a semantic network, is CYC, a very large database of common sense assertions linked together by relations and designed to produce coherent inferences (Lenat and Guha, 1990). CYC is being developed with the hopes of produc-ing a common sense reasoning component for a full artificial intelligence system (see http://www.cyc.com/).

CONCLUSION

In this chapter, we have provided a detailed description of Quillian's original seman- tic memory model and its transition into a somewhat diverse set of generic elements used in subsequent theoretical developments. We have also provided an overview of neuro-re- lated theories of semantic memory, which have incorporated some aspects of the generic semantic models, but which, until the mid-1990s, were driven mainly by findings from particular neuropsychological populations. Over the past 10 years, the neuro-related models of semantic memory have become increasingly informed by functional neu- roimaging findings with both normal and patient populations. Finally, we discussed the storage versus access issue in neuropsychological populations, and then focused on find- ings from two specific paradigms that we and others have used extensively to evaluate the status of semantic memory in AD: similarity judgments (for items within a given semantic domain) and semantic priming. Here, we conclude the chapter with: (1) an attempt to reconcile the present status of semantic memory research with its origins; and some observations regarding the need for further communication and collaboration between cognitive scientists (including, of course, psycholinguists) who are developing models based solely on normal behavioral data, and neuropsychologists or cognitive neu- roscientists who are developing models based on behavioral data from brain-impaired populations and, increasingly, functional imaging data from normal as well as brain- impaired populations.

[This page is intentionally left Blank]

CHAPTER 10
PROSODY

INTRODUCTION

Prosody has been called 'the organizational of speech' (Beckman 1996). From Oxford Dictionary, Prosody is the patterns of sounds and rhythms in poetry, when we talk about literature and writing but when we talked about language, Prosody is the part of phonetics that deals with stress and intonation as opposed to individual speech sounds. Prosodic elements such as stress and intonation are generally seen as providing both 'natural' and properly linguistic input to utterance comprehension. They contribute not only to overt communication but to more covert or accidental forms of information transmission. They typically create impressions, convey information about emotions or attitudes, or alter the salience of linguistically-possible interpretations rather than conveying distinct propositions or concepts in their own right. This paper will be focus on three main issues. First is intonation, second is stress and the third is phrasing.

The goal of this chapter is to review selected literature on prosodic processing in adult speakers and listeners, with a focus on prosodic phrasing and sentence-level prominence. The structure of the chapter is as follows: First, we give a brief history of some early attempts to understand the impact of prosodic structure on comprehension, highlighting some of the basic issues that continue to motivate research in the area and presenting some of the methodological challenges inherent in the study of prosodic effects. Next, we review briefly phonetic measures and phonological annotation methods commonly used to describe the prosody of production data and spoken experimental materials. We discuss strategies for the construction of experimental materials and argue that the precise specification of prosodic form is critical to the understanding and replication of prosodic effects in processing. We then present current research findings on prosodic phrasing and its relationship to syntactic processing, and prosodic prominence as it affects information structure, and interacts with phrasing. For reasons of space, we limit our discussion primarily to research conducted in English, although we note that a thorough and general understanding of the nature of prosodic processing cannot rely on work in a single language. Finally, we review some recent sentence-processing models that incorporate prosodic structure. THEORIES OF PROSODY

Ancient critics like Aristotle and Horace insisted that certain metres were natural to the specific poetic genres; thus, Aristotle (in the Poetics) noted, “Nature herself, as we have said, teaches the choice of the proper measure.” In epic verse the poet should use the heroic measure (dactylic hexameter) because this metre most effectively represents or imitates such qualities as grandeur, dignity, and high passion. Horace narrowed the theory of metrical decorum, making the choice of metre prescriptive; only an ill-bred and ignorant poet would treat comic material in metres appropriate to tragedy. Horace prepared the way for the legalisms of the Renaissance theorists who were quite willing to inform practicing poets that they used “feete without joyntes,” in the words of Roger Ascham, Queen Elizabeth’s tutor, and should use the quantitative metres of Classical prosody. Prosody is a type of distinguishing class (a class that distinguishes language units from other language units) which includes pitch, pressure, and accent.

1. The Middle Ages

During the Middle Ages little of importance was added to actual prosodic theory. In poetic practice, however, crucial developments were to have important ramifications for later theorists. From about the second half of the 6th century to the end of the 8th century, Latin verse was written that no longer observed the rules of quantity but was clearly structured on accentual and syllabic bases. This change was aided by the invention of the musical sequence; it became necessary to fit a musical phrase to a fixed number of syllables, and the older, highly complex system of quantitative prosody could not be adapted to simple melodies that must be sung in sequential patterns. In the musical sequence lies the origin of the modern lyric form.

2. The Renaissance

Renaissance prosodic theory had to face the fact of an accomplished poetry in the vernacular that was not written in metres determined by “rules” handed down from the practice of Homer and Virgil. Nevertheless, the classicizing theorists of the 16th century made a determined attempt to explain existing poetry by the rules of short and long and to draft “laws” by which modern verse might move in Classical metres. Roger Ascham, in The

Scholemaster (1570), attacked “the Gothic…barbarous and rude Rhyming” of the early Tudor poets. He admitted that Henry Howard, earl of Surrey, did passably well as a poet but complained that Surrey did not understand “perfite and trewe versifying”; that is, Surrey did not compose his English verses according to the principles of Latin and Greek quantitative prosody.

3. The 18th century

In 18th-century theory the doctrine of imitation was joined to numerous strictures on “smoothness,” or metrical regularity. Theorists advocated a rigid regularity; minor poets composed in a strictly regular syllable-stress verse devoid of expressive variations. This regularity itself expressed the rationalism of the period. The prevailing dogmas on regularity made it impossible for Samuel Johnson to hear the beauties of Milton’s versification; he characterized the metrically subtle lines of “Lycidas” as “harsh” and without concern for “numbers.” Certain crosscurrents of metrical opinion in the 18th century, however, moved toward new theoretical stances. Joshua Steele’s Prosodia Rationalis (1779) is an early attempt to scan English verse by means of musical notation. (A later attempt was made by the American poet Sidney Lanier in his Science of English Verse, 1880.) Steele’s method is highly personal, depending on an idiosyncratic assigning of such musical qualities as pitch and duration to syllabic values; but he recognized that a prosodic theory must take into account not merely metre but “all properties or accidents belonging to language.” His work foreshadows the current concerns of the structural linguists who attempt an analysis of the entire range of acoustic elements contributing to prosodic effect. Steele is also the first “timer” among metrists; that is, he bases his scansions on musical pulse and claims that English verse moves in either common or triple time. Twentieth-century critics of musical scanners pointed out that musical scansion constitutes a performance, not an analysis of the metre, that it allows arbitrary readings, and that it levels out distinctions between poets and schools of poetry.

4. The 20th century and beyond

After 1900 the study of prosody emerged as an important and respectable part of literary study. George Saintsbury published his great History of English Prosody during the years 1906–10. Sometime later, a number of linguists and aestheticians turned their attention to prosodic structure and the nature of poetic rhythm. Graphic prosody (the traditional syllable and foot scansion of syllable-stress metre) was placed on a securer theoretical footing. A number of prosodists, taking their lead from the work of Joshua Steele and Sidney Lanier, attempted to use musical notation to scan English verse. For the convenience of synoptic discussion, prosodic theorists are sometimes divided into four groups: the linguists who examine verse rhythm as a function of phonetic structures; the aestheticians who examine the psychological effects, the formal properties, and the phenomenology of rhythm; the musical scanners, or "timers," who try to adapt the procedures of musical notation to metrical analysis; and the traditionalists who rely on the graphic description of syllable and stress to uncover metrical paradigms. It is necessary to point out that only the traditionalists concern themselves specifically with metrical form; aestheticians, linguists, and timers all examine prosody in its larger dimensions.

5. Asian theories

The metres of the verse of ancient India were constructed on a quantitative basis. A system of long and short syllables, as in Greek, determined the variety of complicated metrical forms that are found in poetry of post-Vedic times—that is, after the 5th century BCE.

Chinese prosody is based on the intricate tonal system of Chinese languages. In the Tang dynasty (618–907 CE) the metrical system for classical verse was fixed. The various tones of the language were subsumed under two large groups, even tones and oblique tones. Patterned arrangements of tones and the use of pauses, or caesuras, along with rhyme determine the Chinese prosodic forms.

Japanese poetry is without rhyme or marked metrical structure; it is purely syllabic. The two main forms of syllabic verses are the tanka and the haiku. Tanka is written in a stanza of 31 syllables that are divided into alternating lines of five and seven syllables. Haiku is an extremely concentrated form of only 17 syllables. Longer poems of 40 to 50 lines are also written; however, alternate lines must contain either five or seven syllables. The haiku form has been adapted to English verse and is a popular form. Other experimenters in English syllabic verse show the influence of Japanese prosody. Syllabic metre in English, however, is limited in its rhythmic effects; it is incapable of expressing the range of feeling that is available in the traditional stress and syllable-stress metres.

A. Pitch

From an acoustic phonetic standpoint all sounds are air vibrations, and the higher the frequency of the vibrations (typically calculated by noise), the higher the sound tone. The language sound tones that are most easily captured by the hearing instrument are the sound tones produced by the formation of a narrow groove between the vocal cords, and the frequency of air vibrations that they cause is determined by the vibrational frequency of the vocal cords.

One variation of the pitch that accompanies the whole sentence, or part of the sentence, is intonation or melody song. You can hear easily that we don't use the same tone when we speak; just try saying a few sentences on tone one of the guitar, or piano, and soon you hear how strange such a pronunciation is. So almost every word (and it's not uncommon to say just one word) in another sentence is spoken in another tone. From that in the sentence produced a song: intonation.

For the sake of easy analysis of intonation, phoneticians and phonologists use terms such as: "high", "low" (low), "medium" (raid) notes; or high and low are distinguished by numbers alone, e.g. numbers 1 through 4, as in music (but not the same distance between notes in music). For example the sentence “Have you eaten?” The intonation can be analyzed as follows: Intonation starts at 4, decreases to 1 at the time when the word eat, and rises to 3 at the end of the sentence. The exact height can only be determined in phonetic laboratories. However, even without laboratory equipment, we can achieve a satisfying analysis too, especially if we have musical talent, and our ears are "sensitive" to the highs and lows of the tone.

The difference in tone is not absolute, meaning that it is not absolutely necessary for every Indonesian speaker to begin the question sentence at (eg) the frequency of 800 vibrations per second. What is important is that our relative differences can begin to play a song on the piano on which chords we like, and so with the pitch of the sound. The voices of women and men usually differ by one octave in height, but the difference in the relative high and low of certain sentences that he said seems not too much. In fact, the same person can have a lower frequency at one time, can be higher at another time, in both cases the relative difference of certain intonation tones seems not too far apart. The relative differences between the tones used in intonation can differ between languages (e.g. in French at the highest and lowest not too far apart, compared to English), as well as between dialects, and in between speakers of individual languages.

1. Stress and Accent

Stress and accents are difficult to distinguish. The difficulty lies in terms of the terms, and in the facts named by those terms. In other words, the difficulty for some is merely terminological ("terminology" = terminology) and for some it is factual, which is to say facts. The terminological difficulties have not been solved by phonetics and phonologists. The English term stress is often used as the name of an accent, so stress and accent are the same? For example, the French rule that in every word the last syllabus is stressed can also be formulated using the term accent or, for example, the last syllabus before the –ic or –ical suffix must be stressed or also formulated by stating that the syllabus must be given an accent. Let's decide now for such symptoms we use the term accent, not the term stress, the term "accent", not the term "pressure". Don't confuse the meaning of the scientific term "accent" with the meaning of "accent" in everyday language, where "accent" often means "accent" or pronunciation according to a particular language or dialect.

We use the term "pressure" for what is called "amplitude" in natural science (from the Latin word amplitude ‘width’). Amplitude is the "width" of air vibrations. The high frequency is neutral with respect to the amplitude of each vibration. E.g. if you move your hand in front of your chest from the right on the left and back twice a second, then the "frequency" as high as "2" can be carried out with an "amplitude" of ten centimeters or with an amplitude of thirty centimeters. Indeed, a high frequency of 2 seconds you cannot hear. But you can hear sounds that sound in accordance with the frequency you produce. If so, what is the result of the amplitude of the movement? It will be louder if the amplitude is thirty inches, finer if only ten inches. So the amplitude is thirty inches, finer if only ten inches. So amplitude determines the loudness or strength of the sound produced. It is free from frequency.

Here is another terminological difficulty. Because phonetic books often distinguish pressure (stress, and it is relied upon here that stress relates to the magnitude of the amplitude) of "high pressure" and "low pressure" (high stress and low stress). If the terms "high" and "low" aren't misinterpreted, we don't mind. In fact, the danger of misinterpretation exists: the terms "high" and "low" can be considered to be related to the frequency and other frequencies of the amplitude. Therefore let us distinguish, if necessary, stress (stress) as "strong pressure" and "weak pressure" (strong stress and weak stress) the terms "strong" and "weak" are not so misleading, because the magnitude of the vibration amplitude does determine strong sound produced. Pressure, as with tone, is relative, not absolute. If a portion of a speech is spoken in a voice that is stronger than the sound in the other parts of the speech, the relative difference is adequate. The absolute size of sound strength (i.e. with "decibels") is not important for phonetics.

Now let's take a concrete example of pressure. E.g. If I say the sentence I want to go to Buru (meaning Buru island), but for whatever reason you hear "to Boro", and wonder why I want to go to Boro, then I can repeat the sentence by stressing loudly "to Buru". The emphasis here, namely the pronunciation of the words, is speech with greater amplitude; the pressure here is called "contrast pressure": [.....] to Buru, not to Boro.

Now let's move to the accent problem, e.g. with the example above about the accent at the end in French, or the accent on the syllabus before the -ic or -ical suffix in English. Here it is clear that the accent is not used to contrast contrastively the syllable accented from the other syllabuses in the word concerned. So the "lexical" accent is only a structured characteristic of the sound of the word concerned. But apparently you are asking here whether the accent is concretely different from the pressure, meaning whether the accent is possible without greater amplitude. The answer is that there are indeed two types of accents: accents that are carried out with pressure (called "pressure accents", English stress accents, often also called "dynamic accents"); and accents that occur with tones (called "tone accents", English pitch accents, often also called "musical accents"). Accent is either without pressure or without tone indeed never existed. But because the accent can be with pressure, it can also be with the tone, then the accent is not the same as the pressure, not the same as the tone. Even more difficult, the pressure can consist of a greater amplitude, but is often accompanied by a higher pitch as well. E.g. said earlier [.....] to Buru, not to Boro seems to be spoken in a higher tone in the words of Buru and Boro, but the elevation of the tone did not enter the essence of pressure.

2. Speech Pitch, Production and Perception

In strict terms, pitch is the perceptual correlate of F0, the fundamental frequency or repetition frequency of a sound. One should be aware, however, that rather often the notion "pitch" is used to refer to F0 or the repetition frequency itself. In speech F0 is determined by the rate of vibration of the vocal cords located in the larynx. The physiological and acoustic mechanisms by which F0 is controlled are rather intricate 3 4 and will not be dealt with here. An excellent account of these mechanisms is given in Borden and Harris (1983). Rate of vibration of the vocal cords, and thereby F0, is measured in Hertz (Hz; 1 Hz is 1 cycle per second). The range of F0 for each individual speaker mainly depends on the length and mass of the vocal cords. For males in conversational speech this range is typically between approximately 80 and 200 Hz, for females between approximately 180 and 400 Hz, and for young children this range can be even considerably higher. Within this range each speaker has to a large extent active control over F0: a speaker can choose to speak on a high or a low pitch, and can produce pitch rises and falls. However, many details of the actual course of pitch in speech are not actively controlled by the speaker, but are rather involuntary side-effects of other speech processes, often related to the production of particular speech sounds. For example, other things being equal, high vowels like /i/ and /u/ have a higher intrinsic pitch than low vowels like /a/ (Peterson and Barney, 1952; Ladd and Silverman, 1984; Steele, 1986). In vowels following voiceless consonants the voice pitch starts higher than in vowels following voiced consonants (Ohde, 1984; Silverman, 1986). These involuntary aspects of speech pitch superimpose small perturbations on the course of pitch, and often, in the visual analysis of measured pitch fluctuations in speech utterances, make it difficult to identify those pitch variations that are responsible for the perceived speech melody.

3. Perceptual Equality: Close-Copy Stylizations

It has been assumed above that there are many apparently capricious details in the pitch fluctuations in speech utterances that are not actively controlled by the speaker, but are rather involuntary side-effects of other speech production processes. It was also assumed that such involuntary pitch movements do not contribute to the perceived speech melody. A priori this is a bold assumption, comparable to the assumption that involuntary, segmentally conditioned, variations in speech sound durations are irrelevant to the perceived rhythmical structure of speech. As we will see later, the latter assumption does not hold. But the earlier assumption with respect to pitch fluctuations does hold. This can be shown by using the technique of analysis by-synthesis, replacing the original pitch course of an utterance by an artificial one, using for example an LPC-analysis-resynthesize system (Atal and Hanauer, 1971) or the more recent Pitch Synchronous Overlap and Add method (PSOLA: Hamon, 1988; Charpentier and Moulines, 1989).

A first step in this demonstration is the so-called ‘close-copy stylization’ of pitch in speech utterances, as applied by De Pijper (1983) to British English. Fig. 2 gives the natural, measured F0 curve of an English utterance together with its close-copy stylization. A close-copy stylization is defined as a synthetic approximation of the natural course of pitch, meeting two criteria: it should be perceptually indistinguishable from the original, and it should contain the smallest possible number of straight-line segments with which this perceptual equality can be achieved. Note that the graphical representation of the close-copy stylization continues through the voiceless portions in the utterance. In the actual resynthesize voicing, and therewith pitch, will be suppressed in these voiceless portions.

4. Perceptual Equivalence: towards Standard Pitch Movements

Close-copy stylization is based on perceptual equality. It is only a first step in the description of intonation. If we have someone imitate the intonation or speech melody of an utterance, either with the same words or with different words, or even with no words at all by humming, we obtain a pitch curve that will definitely not be perceptually equal to the original. It will be easy to hear many differences. But yet we, or a panel of native listeners, can hear whether the imitation is successful in conveying the same melodic impression. Apparently, intonation is organized in terms of melodic patterns that are recognizable to native speakers of the language. This calls for a unifying notion different from perceptual equality.

For this other use the term perceptual equivalence. Two different courses of F0 are perceptually equivalent when they are similar to such an extent that one is judged a successful (melodic) imitation of the other (‘t Hart et al., 1990:47). Perceptual equivalence implies that the same speech melody can be recognized in two realizations despite easily noticeable differences, in the same way that the same word can be recognized from different realizations.

5. Combining Pitch Movements: towards a Grammar of Intonation.

Once one has defined an inventory of pitch movements for a particular language, it should be possible to generate sequences of such pitch movements. Such sequences would then constitute acceptable melodic realizations for speech utterances. Trying to do that, one will soon find out that not all possible sequences are acceptable. So, for example, in Dutch an accent-representing rise cannot be followed by another accent representing rise without an intermediate fall. Also, studying the distribution of pitch movements in a corpus of utterances, it may become obvious that some pitch movements belong closer together than others: there appears to be a multilevel hierarchical structure to intonation. If we consider pitch movements themselves to constitute the lowest or first level of description, the second level is that of configurations, and the third that of contours.

A configuration is a close-knit into national unit consisting of one or more consecutive pitch movements, for example a rise followed by a fall or a rise followed by a fall followed by a rise. Generally, constraints on combining pitch movements are much stricter within a configuration than at its boundaries. In their description of Dutch intonation ‘t Hart et al. distinguish Prefix configurations, Root configurations, and Suffix configurations, notions that closely resemble the time-honored notions of Head, Nucleus and Tail in the British impressionistic tradition of intonation studies. Prefix configurations are optional and recursive. They always precede another Prefix or a Root. Root configurations are obligatory and non-recursive: each contour must contain only one and not more than one Root. Suffix configurations are optional and non-recursive. A Suffix always follows a Root.

Pitch contours are defined as lawful sequences of configurations. Each pitch contour extends over a clause (in some loose sense of the word clause, referring to a group of words that the speaker has chosen as belonging together, as having some kind of coherence). This entails that multi-clause sentences have as many contours as there are clauses. Because there are recursive elements in contour formation, the number of contours is unlimited.

Many sequences of Prefix, Root and Suffix appear to be unlawful. Therefore explicit rules are needed to generate the lawful sequences and exclude the unlawful ones. The inventory of pitch movements, their combinations in configurations, plus the set of rules generating the lawful contours, together constitute a grammar of intonation. Ideally, such a grammar of intonation generates all and only the acceptable pitch contours of the language. The predictions by the grammar of Dutch intonation were verified against a corpus of 1500 spontaneous and semi-spontaneous utterances, and found to account for 94 percent of the contours in the corpus (‘t Hart and Collier, 1975).

6. Basic Intonation Patterns

For both British English (Gussenhoven, 1983; Gussenhoven, 1984; Willems et al., 1988) and for Dutch (Collier and Hart, 1972; Collier, 1975) it has been shown that pitch contours can be classified into different families. Pitch contours belonging to the same family are put in the same class by native listeners when these are asked to sort utterances into a limited arbitrary number of subjective melodic categories. For both Dutch and English it appears that class membership is determined by one or more pitch movements belonging to the Root configuration. The pitch contours belonging to the same family are supposed to be manifestations of the same underlying “basic intonation pattern”. In the grammar of intonation each basic intonation pattern can be defined as the family of generation paths that go through the Root configurations that corresponds to that pattern. For both British English and Dutch six such basic intonation patterns can be distinguished, probably carrying different attitudinal and/or emotional connotations. For Russian some ten such basic intonation patterns have been distinguished (Odé, 1986). Most of these basic intonation patterns are used rather infrequently. In Dutch, over 60 % of pitch contour tokens one encounters in everyday speech are realizations of a single basic intonation pattern, the so-called “hat pattern”.

7. Text and Tune

So far intonation has been dealt with here virtually without reference to the sequences of words on which it is superimposed in actual utterances. In order to select a fitting pitch contour for a particular sentence, or sequence of pitch contours in the case of longer sentences, one has at least to know two things about the sentence. One has to know which words are to be provided with a pitch accent on their lexically stressed syllable, and whether, and if so where, boundaries between successive clauses in the sentences are to be made. Once these things are known, acceptable pitch contours can be selected for all clauses from the ones generated by the grammar of intonation. Of course, for each clause or sentence with known accent placements, there still is a variety of different possible pitch contours, and each pitch contour, due to its inherent flexibility with respect to time, can be made to fit a variety of different clauses or sentences.

In normal human speech the speaker determines which words are to be accented and where clause boundaries are to be made according to rules and strategies that will be briefly discussed in section 4 of this chapter. In synthetic speech, for example in textto-speech systems, such rules and strategies have to be approximated by automatic text analysis (Kulas and Rühl, 1985; Carlson and Granström, 1986; Allen, Hunnicutt, and Klatt, 1987; Quené and Kager, 1993; Dirksen and Quené, 1993). Once this is done, appropriate and acceptably sounding pitch contours can be generated automatically and synchronized with the synthetic speech. Generally, in synthetic speech rules for generating pitch contours are limited to pitch contours that are manifestations of a single, neutral sounding, basic intonation pattern, as there is no basis to select between different intonation patterns.

B. Some Communicative Functions of Speech Prosody

What is the use of speech prosody in normal speech communication? In normal written or printed text there is, apart from punctuation and the use of capitals, very little that corresponds to prosodic patterns in speech. Yet many people easily read more words per minute than speakers can speak at their fastest rate. There is, however, a major difference between text and speech. Text is spatially presented, such that much of it is simultaneously present to the reader. Speech is not. At each moment in time the sound of speech is nothing more than a momentary disturbance of air pressure. One moment it is there, the next moment it is gone. Because speech is often listened to in the presence of other sounds, continuously decisions have to be made which successive sounds are to be integrated in the utterance being perceived and which are to be rejected as extraneous.

The fleeting nature of the sound of speech also has the consequence that human perceptual processing of speech draws heavily on human short term memory functions. It is all in the mind. A listener cannot go back to the physical stimulus during processing, because that stimulus has forever vanished in the past. Yet we notice than in normal speech a great many phonemes are very rapidly produced, becoming grossly degraded to the extent that they become unidentifiable without context, or even are completely deleted. We may imagine that if this were not so, speech would become much too slow for the listeners to keep attention focused on the contents of the message. As we learn from the comparison with reading, comprehension can go much faster than speech allows. But the less specified segmental structure is, the more support a listener needs from suprasegmentally, prosodic cues. These cues can differentiate between more important and less important information as coded in accent patterns, and also organize the message in chunks that are easily processed by the listener, at the same time revealing aspects of the linguistic structure of the message. These examples of communicative functions of speech prosody will be briefly described below. The list is not exhaustive. Prosody may to a certain extent also be used to characterize utterances as statements, questions, or exclamations

Prosody has multiple functions in literature. For example, poets incorporate it in matters like syntactic phrasing, word segmentation, sentence, accentuation, stress, and phonological distinctions. Generally, authors use it to produce rhythmic and acoustic effects in poetry as well as prose. However, it has expressive and pragmatic functions, because a certain sentence in a given perspective expresses more than just its linguistic meanings.

An expressive content could be an identity of a speaker, his mood, age, sex, and other extra-linguistic features. Pragmatic function of prosody encompasses the attitude of speaker and listener, and provides relationship between a speaker and his or her discourse. It also reflects different features of a speaker and his utterance, emotional state, form of utterance, presence of sarcasm or irony, and emphasis.

1. Accent Patterns and Their Role in Speech Communication

In the act of speaking, some words are accented by means of an accent-representing pitch movement on their lexically stressed syllable, with some concomitant cues such as some extra loudness and some lengthening of the word. Of course we can establish that a particular word is accented without worrying too much how the accent is realized in the act of speaking: the notion ‘accent’ is abstract with respect to its realization, and as such basically refers to the same thing as the notions 'sentence stress' or 'word group stress' (Chomsky and Halle, 1968; Liberman, 1979; Selkirk, 1984). To show how accents are used in speech communication we need to introduce the notion 'focus', used here as in Ladd (1980), Gussenhoven (1983), Selkirk, (1984) and Baart (1987). A constituent, which can be a single word but also a word group or phrase, can be presented by the speaker as in focus (or +focus) by means of an accent on a single word that we call the prosodic head of the constituent. The position of the prosodic head within each potential constituent can be derived from syntactic structure. The reasons why a particular constituent can be put into focus, and thus receive an accent on its prosodic head, do not seem to be particularly well understood. But one of these reasons appears to be the 'newness' to the listener of the information contained in that constituent.

2. Auditory Continuity and the Separation of Simultaneous Voices

Cherry (1953) addressed himself to the question of how one recognizes what one person is saying when others are speaking at the same time, a phenomenon he referred to as the 'cocktail party effect'. Cherry mentioned as possible facilitating factors directional hearing, visual information, individual differences in voice characteristics and dialect and transitional probabilities. Although his main experiments were directed at directional hearing and transitional probability, he also observed that, when all the above-mentioned factors except transitional probability were eliminated by recording two messages spoken by the same speaker on the same magnetic tape, the result may sound "like a babel", but the messages can still be separated.

Darwin (1975) neatly demonstrated that pitch continuity is an important factor in "voice tracking". He presented listeners simultaneously with two different passages of speech, spoken by the same speaker, either or not switching from one ear to the other and vice versa during presentation. The stimulus material was so constructed that four conditions were obtained, a normal condition with no switch, a semantic change condition, in which pitch was continuous on each ear but the verbal message switched ears in the middle, an intonation change condition where the verbal message was continuous on each ear, but intonation switched ears in the middle, and a semantics and intonation change condition, in which both verbal message and intonation switched ears simultaneously. Listeners were instructed to attend to one ear only. Switching the intonation from one ear to the other caused a high percentage of intrusions of the unattended ear, showing that listeners track a voice in the presence of another voice (and in the absence of directional cues) mainly on the basis of pitch continuity.

From Darwin's experiment it is reasonable to assume that perceptual separation of simultaneous speech messages is easier for messages in different pitch ranges than for messages in the same pitch range, where the listener may inadvertently switch to the other message whenever the two pitches cross. This was shown to be correct by Brokx and Nooteboom (1982), in an experiment with resynthesized speech utterances from a single speaker, with artificially manipulated pitches. There were approximately 20% less word perception errors with different than with the same pitch ranges.

Obviously, intelligibility of speech in the presence of other speech is better when the pitches or pitch ranges of the two competing messages are different than when they are the same. This effect can be related to the phenomenon of "perceptual fusion", occurring whenever two simultaneous sounds have identical pitches, and to "perceptual tracking": whenever the pitches of target and interfering speech cross each other, the listener runs the risk of inadvertently switching his attention from the target to the interfering speech.

3. Prosodic boundaries

A sequence of words like "the queen said the knight is a monster" can be read and spoken in at least two different ways: "the queen, said the knight, is a monster", or "the queen said, the knight is a monster". The ambiguity inherent in this sequence of words is disambiguated in speech by prosodic phrasing, producing either a strong prosodic boundary after "queen" and "knight" or after "said".

Even when no such strong ambiguities are present, nevertheless speakers tend to divide their speech into prosodic phrases. Potential positions for prosodic phrase boundaries can be derived indirectly from syntactic trees, by assigning metrical trees to the syntactic trees, and then applying some simple phrasing rules to the metrical trees (Selkirk, 1984; Nespor and Vogel, 1986; Dirksen and Quené, 1993). Selkirk, and also Nespor and Vogel, distinguish between two types of phrases, I-phrases or intonational phrases, separated by major boundaries, and Phi-phrases or phonological phrases, separated by minor boundaries. Phi-phrases are combined into the hierarchically higher I-phrases. Dirksen and Quené, in the context of a text-to-speech system, attempt to implement some of the ideas of Selkirk and Nespor and Vogel in a set of computational rules. Instead of two hierarchically ordered types of phrases, they assume only one type of phrase boundary which may or may not be realized by a speech pause or final lengthening.

High quality speech without grammatical speech pauses within sentences can be highly intelligible and acceptable. But as soon as speech quality is less than normal, or speech is listened to in noisy conditions, the introduction of grammatical speech pauses can help to maintain intelligibility (Nooteboom, 1985). In general, it can be observed that the contributions of prosody to speech perception becomes more important when the segmental quality of speech or the listening conditions become less favorable.

CONCLUSION

Research on prosody over the past 20 years has broadened our understanding of language comprehension in ways that would not have been possible based solely on the study of written language. Indeed, the research on prosody and processing raises doubts that effects observed in reading necessarily correspond to similar effects in auditory comprehension. The same can be said of the underlying cognitive processes that researchers infer from those effects. We anticipate that one possible shift for the field will come with the advent of improved methodologies for tapping prosodic structures during silent reading.

The objectives of the tutorial are at once broad and restricted. The broad objective is to outline the domain of prosody: discussion, within the semiotic rank framework, of a somewhat eclectic selection of 'must know' topics in prosody from details of phonemic tone, accent in context, phrasal intonation and speech timing to the functionalities of prosody in discourse and prosodic differences between languages and language varieties. The narrow objective is methodological: to show quantitative phonetic measurements and visualizations in the acoustic phonetic domain, in the belief that not only does a picture convey more than a thousand words in the context of a tutorial, but that the precision and limitations of a heuristic measurement-and visualization-oriented approach offer a more fruitful long-term perspective than other approaches.

More generally, processing accounts have a long way to go in terms of explaining how the multiple components in prosodic representation work together to influence various other levels during sentence comprehension. There are also interesting questions regarding to what extent segmental, lexical, syntactic, semantic, and discourse factors may influence the perception of prosody, and how that might come about. Although we have focused on studies that investigate English, successful processing accounts will be those that explain similarities and differences in processing across languages that have similar and different prosodic systems.

[This page is intentionally left Blank]

CHAPTER 11
DISCOURSE COMPREHENSION

INTRODUCTION

This paper outlines a theory of macro-structures within a larger framework of a model of cognitive information processing. We will focus on semantic structures and processes of discourse comprehension; in particular, we will formulate the macro-rules underlying the global interpretation of discourse. Although the focus is on discourse comprehension, the basic principies of macro-processing also are valid for other cognitive domains, such as vision, action, thinking, and problem solving. The notion of macro-structure is introduced as a partial explication of such notions as `schema' or 'plan' as they are currently used in cognitive psychology and artificial intelligence. Semantic macro-structures will be distinguished from other kinds of `schematic' structures of discourse, such as narrative super-structures. Finally, we will elucidate the relationship between macro-structures and fr ames. In linguistics, macro-structures have been postulated in order to account for the 'global meaning' of discourse such as it is intuitively assigned in terms of the 'topic' or 'theme' of a discourse or conversation. The assumption is that these notions cannot be accounted for in terms of current logical, linguistic, and cognitive semantics for isolated sentences or sequences of sentences. In disciplines such as rhetorics and narrative theory, macro-structures may constitute the semantic basis for specific categories and rules. For instance, the setting of a narrative should be defined at a macro-level of analysis and cannot be defined in terms of individual sentences (or their underlying propositions). Similarly, wellknown categories such as `premise' and 'conclusion' in an argument also operate on global structures of discourse. Besides defining the global coherence of discourse, macro-structures also contribute to 'local' coherence at the micro-level of connections between propositions in composite sentences and successive sentences. In this paper, we will show that some fundamental problems of cognitive psychology can be accounted for in terms of macro-processing of complex semantic information. If a discourse is at all long, subjects are unable to process it at a micro-level alone: not only are they unable to store and retrieve such discourse verbatim, but they are also unable to retrieve the constituent propositions. At the same time, if the sequences of sentences can be assigned a macro-structure, they will be recalled much better than scrambled sequences. Thus, comprehension (as well as production) probably takes place at several levels, such that lower-level information is organized, reduced, and represented at higher levels. These processes involve the use of macro-rules; the input to the macro-rules is the micro-structure, and the output is the macro-structure. Macrostructures help to explain the ability to summarize discourse, and in general to use information from discourse for other cognitive tasks, even if the individual propositions of the discourse are no longer accessible. Similarly, macro-structures may provide further insight into the structure and use of frame-like representations of conventional knowledge in discourse comprehension and other cognitive tasks.

A. Discourse Comprehension

We will deal primarily with *semantic* macro-structures, although other kinds of global structures (e.g., narrative and pragmatic ones) may also play a role in the global organization and comprehension of discourse. One possible way to make semantic structures explicit is to use a formal language, such that expressions of this language can be given a suitable interpretation in the formal semantics of that logical language.' This formal semantics may be of the usual model theoretical sort, (e.g., involving notions such as possible worlds). One of the difficulties with other kinds of representations (e.g., graphs) is that they do not yet have an associated formal semantics (cf. Woods, 1975). Micro-structures and Coherence Micro-structures, the sequence of propositions underlying the sequence of sentences of the discourse, constitute the input to the macro-rules. The micro-structures themselves require some preliminary semantic description. We will ' introduce the abstract concept of a *text* to refer to the abstract structure of a discourse. We can characterize the structures of texts at different levels of description.2 At one level of description, a text is simply an ordered sequence of propositions, which under various pragmatic, stylistic, and other constraints is mapped onto a sequence of sentences. Propositions are construed in the usual way, namely, as n-place predicates followed by n arguments which may be bound by quantifiers. Propositions may be modalized by various kinds of operators (tense, knowledge, belief, obligation, etc.), and connectives may be used to make compound propositions.

The semantics of the formal language representing propositions provides recursive truth (or satisfaction) conditions in a constructive way. That is, interpretations of larger units depend on the interpretation of smaller units. Predicates are interpreted as properties or relations and arguments as individuals. Modal operators are interpreted in terms of possible worlds; for example, *It is necessary that p* is true in a world wi iff *p* is true in all possible worlds which are accessible from wi. Sentences have as their referents the facts in possible worlds. Thus, the sentence *Peter is ill* refers to the fact that Peter is ill now, i.e., in the actual possible world at the moment of the utterance. However, unlike classical formal languages, natural language is intensional. That is, we want to assign not only extensional referents, but also meanings. In particular, the extensions of an expression are assigned on the basis of its intensional meaning. Such intensions are functions, taking possible worlds as arguments and extensions as values. Thus, the phrase *the book* has an intensional meaning, namely the individual concept of a book, which may take various extensions, i.e., actual books referred to, in particular situations. Both intensional and extensional interpretations are necessary in an account of the semantic structures of discourse.

A semantics of discourse is characterized by relative interpretations: sentences in a discourse sequence are not interpreted in an `absolute' way, but relative to the interpretation of other, mostly previous, sentences of the discourse. Sequences satisfying the constraints of relative interpretation are called *linearly coherent.* One major coherence constraint is connection. Two propositions are considered to be pairwise connected if the facts they denote are related. This relation can be made explicit in terms of possible, probable, or necessary conditions, components, or consequences. Connections between propositions are typically expressed by natural connectives such as *and, because, yet, so,* etc. (cf. van Dijk, 1977). Sequences of propositions exhibit other coherence properties besides pairwise connections. For example, two expressions may both refer to the same facts, properties, or individuals. It should be noted, however, that in general these conditions are neither necessary nor sufficient for coherence. Discourse coherence is not primarily a matter of meaning, but of reference. Roughly speaking, the coherence of the discourse depends on the coherence of the possible-world fragment or course of events it represents. One of the conditions that makes a text uninterpretable is the impossibility of imagining situations where it could be satisfied. Of course, numerous formal and empirical details are omitted in this presentation. A further note of caution ís that our theoretícal abstractions and generalizations apply to an idealized discourse.

Actual discourses that are produced, understood, and accepted do not always have a fully correct textual structure. Additional rules of pragmatics, cognitive strategies, and social conventions must account for the relevant conditions of acceptability of non-ideal discourses. There is another constraint on linear coherence that is not referential, but rather intensional or conceptual; not only must denoted facts be related, but this relation must be relative to a *topic of discourse.* That is, the facts must originate in the same range of semantic space. Thus, John's playing the piano may, as such, be independent from Mary's knitting, but both actions lie in the same range, namely, human leisure activities. Thus, connection conditions can be established relative to other propositions, which may or may not be explicitly stated in the text. For example, a sequence may be coherent because each fact relates to the general theme. The notion of a theme or topic of a discourse or a conversation will be reconsidered in terms of macro-structures. This means that condítíons for linear coherence may depend on condítíons for global coherence.

Coherence is not only semantic, but may also be determined by pragmatic conditions. Clearly, connections between facts should be satisfied not only `objectively', but also relatíve to language users and communicative contexts. Similarly, the connections must relate not only facts but also speech acts. Thus, one speech act may constitute a condition, component, or consequence of another speech act. The details of these various pragmatic conditions on coherence will be ignored here, as well as other principies of cooperative communication and interaction (cf. van Dijk, 1977). A major pragmatic constraint on discourse is that it be informative; consequently, information that the hearer already knows need not be expressed and asserted. This knowledge may be contextual or general. Contextual knowledge pertains to properties of the communicative situation, such as the presence of certain objects. This allows for coherence based on the context, such as the use of indexical pronouns *(1, you, this, that, yesterday, now,* etc.). General knowledge includes lexical/semantic information pertaining to the meaning postulates of the language — for example, knowing that the word "bachelor" entails unmarriedness. Altematively, the general knowledge may be conventional, involving shared knowledge about the `world' (e.g., knowing that Mexico is warm). Conventional knowledge contains not only actual facts, but also `possible facts' compatible with the actual world.

1. The Study of Discourse

a. Historical Background

Several disciplines in the humanities and the social sciences have recently shown an increasing interest in the study of discourse. This development, which really began around 1970, is not without historical sources however. Over 2000 years ago, classical poetics and rhetorics already provided structural models for texts, such as poetry, drama, and legal or political discourse (Wellek, 1955; Wimsatt & Brooks, 1957; Lausberg, I960; Corbett, 1971). The conceptual sophistication of classical rhetorics remained unmatched until the development of structuralism in linguistics, poetics, and anthropology in the late 1960s, after the earlier example of the so-called Russian Formalists (Erlich, 1955) and the Czech Structuralists between the world wars (1hwe, 1972; Culler, 1975). Thus, the work of the Russian anthropologist Vladimir Propp (1928) on the Russian folktale provided an example for the structural approach to narratives which was taken up more than 30 vears later, mainly in France, by anthropologists and literary scholars such as LéviStrauss, Barthes, Bremond, Todorov, Greimas, and others, and which finally emerged within psychology, in the work on story grammars (Rumelhart, 1975; van Dijk, 1980a). Although these various classical and structuralist theories do not meet the current methodological standards of explicitness in linguistics and psychology, many of the theoretical notions remain relevant today.

2. Textlinguistics

Until the 1970s modern linguistics in America rarely looked beyond the sentence boundary. The prevailing generative transformational paradigm focused on phonological, morphological, syntactic, and later also semantic, structures of isolated, context- and text-independent sentences, ignoring the early programmatic call for discourse analysis by Harris (1952). Interest in the linguistic study of discourse was restricted to less prominent linguistic schools, such as tagmemics (Pike, 1967; Grimes, 1975; Longacre, 1976),

which developed discourse analytic methods mainly for descriptive field work on indigenous languages. European linguistics, especially in England and Germany had remained somewhat closer to the structuralist tradition which had less respect for the boundaries of linguistics itself in general, and of the sentence unit in particular (Halliday, 1961; Hartmann, 1964, 1968; Harweg, 1968; Petöfi, 1971; van Dijk, 1972; Dressler, 1972; Schmidt, 1973). Indeed, some of these linguistic studies of discourse were at the boundaries of grammar, stylistics, and poetics (Leech, 1966; Crystal & Davy, 1969). Initially, the more theoretical claims and proposals based on the assumption that a grammar should also account for the systematic linguistic
structures of whole texts, thereby becoming a text grammar, remained in a programmatic stage, still too close to the generative paradigm for comfort. Soon, however, both text grammars and the linguistic study of discourse in general developed a more independent paradigm, finally spreading from Europe to the United States (van Dijk, 1977a; van Dijk & Petöfi, 1977; Dressler, 1978; Sinclair & Coulthard, 1975; Coulthard, 1977, and many other studies; see de Beaugrande & Dressler, 1981, and de Beaugrande, 1980, for a survey and introduction). More or less parallel with this development, American linguistics had itself shown an increasing tendency toward text- and context-dependent grammatical analysis, after the earlier tagmemic work, especially within so-called functional paradigms (Givon, 1979a).

3. The Social Sciences and Discourse Analysis

The study of discourse became relevant in particular as soon as it was recognized, also around 1970, that language studies should not be restricted to the grammatical analysis of abstract or ideal language systems, but, rather, that actual language use in the social context should be the empirical object of linguistic theories. Thus, sociolinguistics not only became interested in the study of social variation of language use, but also paid increasing attention to various forms of language use, such as verbal dueling and storytelling (Labov, 1972a, 1972b). Some of this sociolinguistic work became intertwined with a similar developmentin anthropology and ethnography, where earlier structural analyses of myths, folktales, riddles, and other forms of verbal art gave way to a broader analysis of communicative events in various cultures (Gumperz & Hymes, 1972; Bauman & Sherzer, 1974; Sanches & Blount, 1975). Finally, this general tendency toward a study of naturally occurring speech could also be observed in microsociology, where the ethnomethodologicat attention paid to everyday interaction soon focused on conversational interactionas well (Sudnow, 1972; Sacks, Schegloff, & Jefferson, 1974; Schenkein, 1978). In fact, conversation analysis soon became so popular that it virtually was identified with discourse analysis, and its influence in recent linguistics has been considerable (Franck, 1980; Coulthard & Montgomery, 1981). At the moment, it is hard to make strict disciplinary distinctions within the study of discourse, which seems to emerge more and more as an independent, interdisciplinary field, in which purely linguistic or grammatical methods and theories mingle with those from ethnography, microsociology and, as we will see, from psychology.

4. Psychology and Artificial Intelligence

Following the prevailing generative transformational trend, psychology and psycholinguistics were hesitant to recognize the relevance of discourse to the study of language processing. Early psycholinguistic models in the 1960s were restricted the syntax and, later, the semantics, of isolated sentences (Clark & Clark. 1977; Fodor, Bever, & Garrett, 1974). Again, the early 1970s brought a breach in this paradigm. The growing interest in semantic memory resulted in the use of discourse materials and the first steps toward a cognitive model of discourse understanding (Kintsch, 1972. 1974; Bower, 1974; for a survey of other work, see Chapter 2). At the same time, educational psychology realized that learning often takes place on the basis of texts, which also contributed to the quickly developing interest in memory for discourse (Rothkopf, 1972, Meyer, 1975). Thus, we witnessed in psychology a general revival of earlier work on discourse within the gestalt tradition, notably that of Bartlett (1932), which had only occasionally inspired psychologists during the intervening 40 years (Cofer, 1941; Gomulicki, 1956; Paul, 1959; Slamecka, 1959; Pompi & Lachman, 1967). This revival, not only of discourse comprehension but also of various schema theories, took place in artificial intelligence as well. In this area, the year 1972 brought a decisive paradigm shift (Winograd, 1972; Charniak, 1972; Simmons, 1972). The computer-simulated understanding of language required the development of programs for the automatic processing of texts. Crucial to this research was the modeling of world knowledge necessary for the understanding of stories, for example. Thus, Bartlett’s notion of schema was taken up again in a more explicit fashion under such labels as “schema,” “scenario,” “frame,” and “script,” in order to account for the role of world knowledge representations in discourse understanding and other complex cognitive tasks (Schank & Colby, 1973; Minsky, 1975; Bobrow & Collins, 1975; Norman & Rumelhart, 1975; Schank & Abelson, 1977). Ten years have elapsed since these early studies in linguistics, psychology, artificial intelligence, and the social sciences.Whereas the earlier approaches often developed in a more or less autonomous and parallel way, we now witness an increasing integration of the many theoretical proposals. Within the wide new field of cognitive science, the interdisciplinary study of discourse has seen the publication of numerous books and papers, the foundation of two specialized journals (*Discourse Processing*, 1978; *Text*, 1981), and the regular occurrence of conferences or sessions within larger conferences. There have been numerous mutual contacts between linguistics and psychology, between linguistics and microsociology, and between psychology and ethnography. In our initial work on cognitive models for discourse comprehension (Kintsch & van Dijk, 1975, 1978; van Dijk & Kintsch, 1978) we attempted to integrate several proposals from these earlier approaches to discourse, in particular from our own work in these areas. Thus, the general memory model was developed from previous work on semantic memory (Kintsch, 1970, 1972), whereas the various textual structures, such as local and global coherence, macrostructures, and superstructures, were analyzed for their role in processing in terms of earlier textlinguistic work (van Dijk, 1972, 1977a) and its influence in psychology (Kintsch, 1974). Although this interdisciplinarily inspired model of discourse comprehension has been steadily extended and refined over the past years, both by ourselves and, often independently and into other directions, by others (see the survey of this work in Chapter 2), the model presented in this book should be considered both as a further extension of this earlier work as well as a new direction in the cognitive modeling of discourse processing. Whereas our earlier model could still be characterized as predominantly *structural*, we now propose a more dynamic, processoriented, on-line model, an approach we want to call *strategical*.

B. Basic Assumptions

Having sketched some of the historical background of our model, we shall now present an informal outline of its basic assumptions. These assumptions not only inspire the major theoretical notions and components of the model, but also establish the necessary relationships with other models of discourse use in linguistics and the social sciences. In the next main section of this chapter we give an overview of the major components of the model, which the following chapters will systematically treat in further theoretical and experimental detail.

1. Cognitive Assumptions

Suppose someone witnesses a car accident. We assume that such a person constructs a mental representation of that accident, and that his or her understanding of the observed events consists in that process of construction and its memorial Consequences, Now, suppose that another person hears a story about the same accident. We assume that understanding such a story also involves the construction of a mental representation of the story. Of course, a representation of the accident Itself a representation of the story about the accident will not be identical. In the latter case, we will have a representation of the speaker's already coded version of the accident (Hörmann, 1976). But, the common characteristic of both cognitive is that the person who witnesses the accident and the person who listens to the story each constructs a representation in memory, on the basis of the visual and the linguistic data, respectively. We will call this the *constructivist* assumption of our model. Next, we will assume that both the witness of the accident and the listener of the accident story do not merely represent the visual and the verbal data, such as the movements of objects or persons (events) or the sounds uttered when the story is told, but also, or rather, an interpretation of the events and the utterance (Loftus, 1979). In both cases they construct a meaning: The events are interpreted as 'an accident' and the story utterance is interpreted as a story about an accident. We will call this the *interpretative* assumption of the model. In fact, we will be nearly exclusively dealing with this semantic aspect of discourse processing. We will further assume that the construction of a representation of the accident or the accident story, and in particular of the meaning of the input data, takes place more or less at the same time as the processing of the input data. In other words, we assume that the witness and the listener in our example do not first and store all input data of the respective events. and only afterward try to assign meaning to these. That is, understanding takes place on-line with the processing of input data, gradually, and not post hoc. Using the computer metaphor, we will call this the on-line assumption of discourse processing (Marslen-Wilson & Tyler 1980). Persons who understand real events or speech events are able to construct a mental representation, and especially a meaningful representation, only if they have more general knowledge about such events. In order to interpret some events as an accident, they must know something about the usual traffic events and actions in oich cars and drivers are involved, and for stories they must have more general ledge about stories and about their relationship to the events that they tell of. Similarly, the two persons may interpret the events in the light of previous experiences with similar events, experiences that may have led to the more general knowledge about them. In addition to this knowledge, the listener and the witness may have other cognitive information, such as beliefs, opinions, or attitudes regarding auch events in general, or motivations, goals, or specific tasks in the processing of these events. More generally, then, we will assume that understanding involves only the processing and interpretation of external data, but also the activation and use of internal, cognitive, information. Since this information can be considered as cognitive presuppositions of the construction process, we will call this the *presuppositional* assumption of the model. As we will see in somewhat more detail in what follows, accidents and stories will not simply be observed and understood in rucmo, hut as parts of more complex situations or social contexts. Understanding them therefore also means that the person uses or constructs information about the relationships between the events and their situations. That is, the understander now has three kinds of data, namely, information from the events themselves, information from the situation or context, and information from the cognitive presuppositions. This information may be combined in an effective way, such that a mental representation of the event is constructed as soon as possible and as well (as meaningfully, usefully, etc.) as possible. This may mean, for instance, that the observer of an accident even constructs meanings derived from his or her presuppositional information for which the external data are lacking, and the same is true for the listener of the story: He or she may have expectations about what may be told before actually having heard it, and this may facilitate the understanding process when he or she actually does get the relevant external information. There is no fixed order, at each point, between input data and their interpretation: Interpretations may be constructed and only later matched with input data. We see that persons have the ability to flexibly make use of various kinds of information, that the information may be processed in several possible orders, that the information that is interpreted can be incomplete, and that the overall goal of-the process is to be as effective as possible in the construction of the mental representation. We will call this the *strategic* assumption of the model. Whereas the other assumptions have already received due attention in previous discourse-processino models, this strategic assumption will be the focus of the present book. We will see that it is inextricably linked with the other assumptions, especially With the on-line assumption about complex information processing of events and discourses. We can now conclude that the major dimensions of our model are based on the assumption that discourse processing, just like other complex information processing, is a strategic process in which a mental representation is constructed of the discourse in memory, using both external and internal types of information, with the goal of interpreting (understanding) the discourse. Of course, these very general assumptions have many corollaries and implications. Thus, the constructivist assumption has as an important corollary that gradual, on-line, construction is possible only on the basis of a structural analysis and synthesis process, in which, at various levels, meaningful units can be distinguished, as can wavs in which these units can be combined into more complex units. This and other corollaries and implications of our assumptions will be spelled out in the appropriate chapters of this book.

2. Contextual Assumptions

We already suggested that discourses such as stories do not occur ill ruc-uo. They are produced and received, by speakers and listeners, in specific situations within a wider sociocultural context. Hence, discourse processing is not merely a cognitive event, but also a social event. This is obvious, of course, but here we will assume, first of all, that the social dimensions of discourse interact with the cognitive ones. In other words, the cognitive model should also provide for the fact that discourse, and hence the process of understanding a discourse, is functional in the social context. We will call this the (social) *functionality* assumption. The first cognitive implication of this assumption is that language users construct a representian not only of the text but also of the social context, and that these two represenaions interact. More specifically, we assume that a story about an accident is told and understood in a process of communication, in which a listener acquires information from the speaker, in this case about some accident (and about the way the speaker has coded this accident in his or her memory). This communicative assumption may mean, among other things, that the listener does not merely attempt to construct his or her own representation of the story, but matches this interpretation with a representation of the assumptions about what the speaker intended the listener to understand. Because intentions are involved in discourse, we deal not only with linguistic objects, but also with the results of some form of social action.

Thus, when telling a story a speaker will engage in the social act, a speech act, of asserting something, or Amning the listener about something. The form and the interpretation of the story wuv be a function of this intended speech act function of the utterance act. We will call this the *pragmatic* assumption of a model of discourse processing. The cognitive implication of this assumption is, for instance, that a person who interprets a story will also construct a representation of the possible speech acts involved, by msigning a specific function or action category to the discourse utterance, and hence to the speaker. In this case, the listener will evaluate the discourse on a number of points relative to the intended pragmatic functions: This story may be pragmatically appropriate as a speech act only if some contextual conditions match with some textual properties. Next, it should be assumed that the interpretation of a discourse as a specific speech act (or series of speech acts) is embedded within an interpretation of the whole interaction process taking place between the speech participants. Both the speaker and the listener will have motivations, purposes, or intentions when engaging in verbal interaction, and the same holds for the further actions with which the verbal actions are related in the same situation. Hence, the pragmatic assumption should be generalized to an *interactionist* assumption. Again, this means that we ussume that language users construct a cognitive representation of the verbal and nonverbal interaction taking place in the situation.

This would imply, for instance, that the representation of the discourse in memory will depend on the assumptions of the listener about the purposes (goals) and further underlying motivations of the ,peaker, as well as on the listener's own goals and motivations when listening to a story. Finally, as we have already suggested, the interaction in which the processing of discourse is embedded is itself part of a social situation. The speech participants may have certain functions or roles; there may be differences in location or setting; and there may be specific rules, conventions, or strategies governing possible interactions in such a situation. One cannot say just anything in any situation: Possible actions, hence possible goals and hence possible discourses, are con strained by the various dimensions of the situation. The accident story may be told in a bar, to a friend at home, or perhaps to a stranger on the bus, but would not be a permissible speech act during an exam. In order to be able to understand a story.

Therefore, we have to link its pragmatic function with the generalintcractional constraints as determined by, or as determining, the social situation; and this is possible only if, again, we specify in our model how the social situation is cognitively represented. In more concrete terms: The interpretation of the meaning and the functions of the accident story will be different when told in informal contexts to our friends than when told, by a witness, in a court trial related to the accident. Hence, we will ultimately have to take into account a .simaiunnl assumption about discourse processing. This may include, as presuppositions, general norms and values, attitudes, and conventions about the participants and the interactions ill some situation. It goes without saying that these various contextual assumptions about discourse processing can be independently formulated within sociological models of language use. Yet, our general functional assumption suggests that the process of' understanding also involves these various kinds of contextual information, that representations are constructed of the speech act, the communicative interactions, and the whole situation, and that these representations will strategically interact with the understanding of the discourse itself. Hence, understanding is no longer a mere passive construction of a representation of a verbal object, but part of an interactive process in which a listener interprets, actively, the actions of a speaker. It will not be our main task to investigate the nature of the representations and the interpretation processes of such contextual information, but we will take them into account when formulating the processes of discourse understanding.

The field of discourse processing has grown tremendously during the past decade, which has resulted in several new journals, societies, and conferences. It is beyond the scope of this chapter to provide a comprehensive coverage of the exciting new empirical findings and theoretical developments. We focus here on the comprehension of written text. The scenario to imagine is a college student reading a literary short story for enjoyment, or studying a technical text for an examination. Thus, the emphasis is on written text rather than oral conversation, and on comprehension rather than the production of discourse. We also focus on the representation of meaning, which includes semantics, pragmatics, and the body of knowledge conveyed in the text. The more shallow levels of code (such as phonology, intonation, syntax, and the lexicon of word meanings) are addressed only to the extent that they help clarify how meaning representations are constructed. We recommend Gernsbacher's (1994) Handbook of Psycholinguistics for readers who desire a comprehensive coverage of psycholinguistics and all levels of discourse processing. An excellent coverage of oral discourse is provided in books by Clark (1993), Levelt (1989), and Rubin (1995). Discourse psychologists have investigated a broad array of written texts. At one extreme, researchers investigate naturalistic texts that are written by professional writers for the general public (van Oostendorp & Zwaan 1994). In the narrative genre, the texts have ranged from simple well-formed folktales to literary short stories (Dixon et al 1993, Gerrig 1993, Kreuz & MacNealy 1996, Miall & Kuiken 1994). Texts in the expository genre have frequently covered topics in history (Perfetti et al 1995, Voss & Silfies 1996) and science (Chi et al 1994, Kintsch 1994).

These investigations of naturalistic text uncover a representative set of discourse features, patterns, devices, meanings, and comprehension processes that are prevalent in a culture. However, the advantage of ecological validity comes at the cost of losing precise control over the texts' stimulus properties. Consequently, at the other extreme, experimenters carefully craft texts to manipulate independent variables, control for extraneous variables, and satisfy counterbalancing constraints. We call these experimenter-generated materials "textoids" because they are not naturalistic discourse segments that are written to convey an informative or interesting message to a comprehender. Indeed, the texts in far too many experiments are meandering, choppy, pointless, and uninteresting; such texts may impose discourse comprehension. There is the risk that the study of textoids unveils unnatural representations and processing strategies. Discourse psychologists are on solid footing when a hypothesis is confirmed in a sample of naturalistic texts in addition to properly controlled textoids. The methods of investigating text comprehension are quite diverse (Haberlandt 1994). Sometimes the objective is to study the meaning representations that are established after comprehension is completed. Claims about these mental representations are tested by collecting recall protocols, summary protocols, answers to questions, and various judgments on test statements. However, these "off-line" measures are not well suited to capturing the processes and representations that are constructed "on-line" during comprehension. What measures and tasks uncover on-line comprehension processes? This question has been debated at length and is far from settled. One straightforward approach is to collect reading times as readers normally read the text. In eye tracking experiments, the researcher records gaze durations on individual words and patterns of eye movements across the words (Garrod et al 1994, Just & Carpenter 1992, Rayner et al 1994). Self-paced reading times are collected by having the reader press a response key after reading individual text segments, such as words, clauses, sentences, or paragraphs (Haberlandt & Graesser 1985). Although these reading times are natural, the times can sometimes be ambiguous with respect to the contents and types of processes they index. Additional clarity is provided in tasks that periodically interrupt the reader and collect data during comprehension. For example, in a "think aloud" task, the reader expresses ideas that come to mind as each clause in the text is comprehended. The content extracted from think aloud protocols is a very rich source of data for discovering possible comprehension strategies and for testing detailed claims about the representations that enter the reader's consciousness.

Researchers have also demonstrated that think aloud protocols are somewhat valid reflections of normal comprehension activities (Chi et al 1994, Ericsson & Simon 1993, Trabasso & Magliano 1996, Zwaan & Brown 1996). However, the protocols do not reliably tap unconscious comprehension processes. Both conscious and unconscious comprehension processes can be tapped in a word-naming task in which readers are periodically interrupted during comprehension and asked to name a test word as quickly as possible. The word-naming latency should be quick if the features of the word closely match a representation that is active in the reader's mind. As an alternative to the word-naming latencies, researchers frequently collect lexical decision latencies on test strings (i.e. whether a sequence of letters forms a word or a nonword), or word recognition latencies (i.e. whether a test word appeared earlier in the text). Unfortunately, there is a drawback to these tasks that interrupt the reader for data collection: The reader might suffer from "comprehension interruptus" and resort to constructing an unnatural, choppy, shallow representation. Therefore, the rigorous discourse psychologist insists on converging evidence from multiple methods before accepting an empirical claim as valid. Some researchers have advocated a "three-pronged method" that coordinates (a) predictions generated by theories, models, and hypotheses; (b) data from think aloud protocols; and (c) behavioral measures that assess processing time (Graesser et al 1994, Magliano & Graesser 1991, Millis & Graesser 1994, Suh & Trabasso 1993, Zwaan & Brown 1996)

C. Multiple Levels of Discourse Representation

Several levels of discourse representation have been identified by scholars in text linguistics, computational linguistics, sociolinguistics, and literary studies. However, some of these levels have not been embraced by discourse psychologists because they are esoteric or are applicable to a very narrow set of discourse contexts. Most discourse psychologists adopt van Dijk & Kintsch's (1983) distinctions among the surface code, the textbase, and the referential situation model. The surface code preserves the exact wording and syntax of clauses. Comprehenders normally retain the surface code of only the most recent clause unless aspects of this surface code have important repercussions on meaning. The textbase contains explicit text propositions in a strippeddown form that preserves meaning, but not the exact wording and syntax. The textbase also includes a small number of inferences that are needed to establish local text coherence. The situation model is the content or the microworld that the text is about. The situation model for a story refers to the people, spatial setting, actions, and events in the mental microworld. This microworld is constructed inferentially through interactions between the explicit text and background world knowledge. In addition to these three levels of representation, psychologists normally acknowledge representations and processes at two other levels, which we call the communication level and the text genre level. The communication level refers to the pragmatic communicative context within which the text is embedded.

Thus, the writer prepares the text to communicate ideas to readers (Nystrand 1986), and story narrators communicate episodes to narratees. Regarding text genre, discourse analysts have identified many categories and subcategories of text genre (Biber 1988), such as narration, exposition, description, persuasion, jokes, and so on. A newspaper article, for example, involves quite different structural components, features, and pragmatic ground rules than a joke. All five of these levels contribute to the meaning representations that readers build during comprehension. Moreover, it is a profound understatement to say that these various levels interact with one another in complex ways that are not well understood. To illustrate the five levels of representation, consider the excerpt below that was extracted from the novel Einstein's Dreams by Alan Lightman (1993, p. 102): A mushy, brown peach is lifted from the garbage and placed on the table to pinken. It pinkens, it turns hard, it is carried in a shopping sack to the grocer's, put on a shelf, removed and crated, returned to the tree with pink blossoms. In this world, time flows backward. The text genre is literary narrative. The excerpt is extracted from the beginning of a chapter, somewhere in the middle of the book. The novel has a series of chapters that describe different fictitious villages in Switzerland in 1905. Each village directly challenges our normal concept of time by transforming a basic assumption in our TIME schema. For example, the citizens in one village know about the future but not the past, which is opposite to one assumption in our TIME schema. In the village described above, time flows backward, which clearly violates the normal forward flow of time, from past to present to future. At the pragmatic communication level, the writer or narrator is attempting to unveil fresh insights about time, reality, and life to the reader by violating the normal assumptions about time.

The writer has used a well-known literary device called defamiliarization (Miall & Kuiken 1994). That is, prototypical concepts are transformed in an unfamiliar way by stylistic devices, which forces the reader to reinterpret referents and view them in a new perspective. The events in the first two sentences are very difficult to comprehend as they are being read on-line because there are no obvious causal connections between successive events. The sequence of events in this situation model is incoherent. Then the third sentence reveals that time flows backward; consequently, the order of events in the explicit text is opposite to the normal flow of events in a generic fruit distribution schema. A diligent reader would have to reinterpret the situation model that was constructed from the first two sentences. It is uncertain at this point exactly what deep messages the author wants to communicate by crafting a text with discrepancies among (a) the presentation order of events in the text, (b) the order of events in a generic fruit distribution schema, and (c) the chronological order of events in the situation model for that village in Switzerland. The textbase is normally represented as a structured set of propositions. A proposition refers to a state, event, or action and may have a truth value with respect to a real or imaginary world. Each proposition contains a predicate (e.g. main verb, adjective, connective) and one or more arguments (e.g. nouns, embedded propositions). Each argument has a functional role, such as agent, patient, object, or location. The textbase of propositions is presented below for the first sentence in the example excerpt. PROP 1: lift (AGENT = X, OBJECT = peach, SOURCE = from garbage) PROP 2: brown (OBJECT = peach) PROP 3: mushy (OBJECT = peach) PROP 4: place (AGENT = X, OBJECT = peach, LOCATION = on table) PROP 5: pinken (OBJECT = peach) PROP 6: [in order] to (PROP 4, PROP 5) PROP 7: and (PROP 1, PROP 4) The seven propositions have predicates that are verbs (lift, place, pinken), adjectives (brown, mushy), and connectives (in order to, and). The arguments include objects (peach, garbage, table), an unidentified agent (X), and embedded propositions (e.g. PROP 4 and PROP 5 are embedded in PROP 6). Note that the propositional textbase does not capture several features of the surface code, such as tense, aspect, voice, and the determinacy of the nouns. For example, the textbase does not capture the fact that the sentence syntax is in the passive voice rather than the active voice. It does not indicate that peach has an indeterminate referring expression (i.e. a peach) whereas table is determinate (i.e. the table).

1. Separation and Interaction of levels

Most researchers believe that the five levels of representation exist and are sufficiently distinct for researchers to isolate. However, these beliefs have been challenged. For example, there is not a perfect consensus that there is a separate textbase. Instead, the syntactic composition and lexical items may directly serve as cues or processing instructions on how to construct the situation model, without there being any intermediate textbase of propositions (Gernsbacher 1990, Givón 1992, Perfetti & Britt 1995). Similarly, the reader of a novel may not construct an invisible, virtual writer or storyteller that communicates with the reader, unless there are explicit features in the text that signal that communication level. Instead, the reader may merely become absorbed in the microworld as a voyeur or side participant (Duchan et al 1995, Gerrig 1993). A persistent challenge has been to devise experimental tasks that isolate the separate levels of representation. Discourse psychologists have collected sentence recognition judgments in an effort to tease apart the surface code, the textbase, and the situation model (Kintsch et al 1990, Schmalhofer & Glavanov 1986, Zwaan 1994). After reading a text, the participants are given a recognition test on the following classes of test sentences: (a) the original sentence verbatim, (b) a paraphrase of the original sentence, (c) a plausible inference with respect to the situation model, and (d) a false statement. A subtraction procedure is used to define the surface code (a minus b), the textbase (b minus c), and the situation model (cminus d). This approach to measuring the three discourse levels has produced theoretically sensible results. For example, there was a rapid decay of the surface code as a function of retention interval and a very slow decay of the situation model, with the textbase in between. When readers believe they are reading literature, the surface code is enhanced, and the situation model is reduced compared with when readers believe they are reading newspaper articles (Zwaan 1994). Therefore, readers are concerned about what is true about the world when they read newspaper articles, whereas they attend to more of the wording and stylistic devices when they read literature. Results such as these suggest that there are natural demarcations among the surface code, the textbase, and the situation model. Kintsch and his associates have also explored individual differences among readers in an effort to segregate differences between the textbase and the situation model (Kintsch 1994, Mannes 1994, McNamara et al 1995). In McNamara et al, a technical text on the functioning of the heart was studied by students who varied in their background knowledge about the heart (low versus high knowledge). The coherence of the textbase was manipulated by having different versions of the text. Text coherence was enhanced by linking clauses with appropriate connectives and/or by inserting topic sentences, headings, and subheadings at appropriate locations. After studying the texts, the students were tested with tasks that tap the textbase (such as recall for the text) and tasks that tap the situation model (such as difficult questions that require reasoning and problem solving). The results for the low-knowledge readers were compatible with virtually all theories of comprehension. That is, a coherent textbase enhanced performance on measures of both the textbase and the situation model. For high-knowledge readers, however, the pattern of results was more interesting. A coherent textbase slightly enhanced recall but actually lowered performance on tasks that tap the situation model. This cross-over interaction supports the claim that the textbase can be separated from the situation model. Moreover, these results have intriguing implications for education and the writing of textbooks. A coherent textbook improves learning for readers with low knowledge, no matter how the learning is measured. However, readers with an adequate background knowledge may actually benefit from a text with coherence gaps and other obstacles that prevent superficial processing. A coherent text that explicitly lays out the material may give readers with comparatively high knowledge an illusory feeling that they have understood all of the explicit text and its implications, when in fact their representations are imperfect at the deeper situation model.

CHAPTER 12
FIGURATIVE LANGUAGE

INTRODUCTION

Figurative language is a language that uses words or expressions with a meaning that is different from the literal interpretation. In addition, figurative language is used in any form of communication, such as in rarely used in daily conversations, opinion in newspaper, advertisements, novels, poems, etc. Moreover, figurative language is the use of words that go beyond their ordinary meaning. It requires the readers to use his/ her imagination to figure out the author's meaning. It makes figurative meaning is difficult to understand because the readers cannot find the meaning of the figurative language in the dictionary just like the other vocabulary words that the readers usually use in our daily conversation. To know the meaning of figurative language the readers need to use his/her imagination to imagine what the words are said or what the words refer to Yuri (2013) studied about figurative language found in an advertisement like personification, hyperbole, simile, and metaphor that tend to be used in internet advertisements. The advertisements were classified into male and female products. These classifications are in order to show types of figurative language, which is dominant and not dominant between male and female products. She concluded that the use of figurative language in advertisement making these advertisements more exciting and interesting, the advertisers used figurative language to make the product become alive. The use of figurative language in advertisement like in internet advertisement, gives big influence in promoting the product. It is easier the producers deliver the message of their product to the consumers.

Further, figurative language is not only added in advertisements but also in news. As source of information or news, the presence of opinion column has a very important role in delivering the opinion to the readers. By reading the opinion, the reader can increase their knowledge; they can follow the progress of issues in the world that is developing rapidly. In conveying the opinions, the journalists used many techniques to make their opinions interesting to read. One of these techniques is the use of what is called figurative language. That is why the journalists apply figurative language in the opinions.

The writer interested to find out the figurative language that is used in local newspaper in West Sumatera called Padang Ekspres newspaper. Padang Ekspres newspaper is one of newspaper that has many readers in West Sumatera. The writer read Padang Ekspres newspaper and found many figurative languages in Padang Ekspres newspaper than other newspaper. In this case, the writer choose opinion column as source of data, because in this column the writer found many figurative languages.

Commonly figurative language found in written language, for example, in the opinion column. In this column, the journalism used figurative languages to persuade the reader to read opinion. The properties are a thing that must be met by a variety of language journalism given the newspaper read by all levels of society are not the same level of knowledge. This is because not everyone interested to read the newspaper. Therefore, the language of journalism is prioritizing aspects that can interest readers, for example, by using figurative language.

Figurative language is using figures of speech to be more effective, persuasive and impactful. Figures of speech such as metaphors, similes, allusions go beyond the literal meanings of the words to give the readers new insights. The journalist used these figurative languages to compare something with another thing. In the other words, they used figurative language to make the sentence more effective and persuasive. Figurative language can appear in multiple forms with the use of different literary and rhetorical devices.

Based on the explanation above, the writer interested to find out the figurative language that used in local newspaper in West Sumatera called Padang Ekspres newspaper. Padang Ekspres newspaper is one of newspaper that has many readers in West Sumatera. The writer found many opinions are frequently written by the journalists in the form of figurative language. Furthermore, the writer thinks that not all of the readers understand the meaning of figurative language in the opinion column.

There are two reasons why the writer chooses the opinion column online Padang Ekspres newspaper. First, opinion column is an interesting column that discusses trending topics or hot issues. Then, there are many figurative languages that can be a source of data from opinion column Padang Ekspres Newspaper.

A. Distinguishing Figurative from Literal Language

One of the continuing difficulties with the psycholinguistics literature on figurative language understanding is that few scholars ever attempt to define the terms “literal” and “figurative.” A traditional assumption in many academic disciplines is that literal meaning is primary and the product of default language comprehension. Thus, in psycholinguistic terms, the human language processor is designed for the analysis of literal meanings. Nonliteral, indirect, and figurative meanings are secondary products, and dependent on some prior analysis of what words and expressions literally mean. This general theory implies that nonliteral meanings should always take more time to interpret than are literal meanings.

Psycholinguistic research over the past 40 years has struggled to create adequate accounts of sentence parsing and discourse processing. Although there has been significant progress in our understanding of different aspects of on-line sentence processing in regard to specific topics (e.g., the interaction of syntax and semantics in sentence parsing, reference assignment, ambiguity resolution, establishing coherence relations in text), there is no single agreed upon position as to what people ordinarily do as they encounter language word by word in speech and reading. Thus, there is really not a single position on literal meaning processing. This state of affairs highlights the absurdity of theories of figurative speech processing that are often based on unverified assumptions as to how so-called literal language is usually understood.

In fact, it is not clear what the operational definition of “literal” meaning is in most psycholinguistic experiments. These studies individually compare metaphoric vs. literal meaning, ironic vs. literal meaning, idiomatic vs. literal meaning, metonymic vs. literal meaning, and so on. But across the vast number of empirical studies that have compared “literal” and “figurative” meaning, the variety of forms for literal utterances is as great as are the differences between metaphors, metonymies, ironies, and so on. Yet scholars continue to assume that the literal meaning they examine empirically somehow is the same variable that other researchers investigate in their respective experiments.

A related tendency in research on figurative language has been to note the difficulty in making a principled distinction between literal and figurative language, or meanings, and to suggest, alternatively, that literal and figurative represent different ends of a continuum of meaning. This idea is seen as especially useful in recognizing that some instances of figurative language, such as novel, poetic metaphor seem more nonliteral than are highly conventionalized phrases which almost seem to express literal meanings (e.g., “kick the bucket” has “to die” as one of its literal meanings). Individual word meanings may alsovary along this literal vs. figurative continuum.

But making these distinctions, even along some graded continuum makes little sense, especially if one is trying to squeeze all aspects of literal and figurative meanings onto a similar scale. Without some consistent idea of what constitutes the notions of “literal” and “figurative” meanings, there is no way of defining the extremes of this proposed continuum. For example, the most novel, poetic instances of metaphor and irony differ from each other in numerous ways (e.g., irony requires meta-representational inferences to be understood in a way that metaphor does not – see Colston & Gibbs, 2002). Even novel metaphors may differ dramatically with some being spectacular instantiations of wellknown conceptual metaphors (e.g., “Our marriage was a roller coaster ride through hell” related to RELATIONSHIPS ARE JOURNEYS) and others reflecting completely new “one-shot” mappings (e.g., “The soul is a rope that ties heaven and earth”). On the literal side of the continuum, different instances of so-called literal meanings may vary along numerous dimensions, depending in part on what aspects of literality are being emphasized (e.g., subject-matter literality, conventional literality, context-free literality, truth-conditional literality) (Gibbs, 1994). For these reasons, the well-intended move toward thinking about literal and figurative meanings as existing along some continuous dimension makes little sense. There is simply no single dimension along which all instances of literal and nonliteral meanings nicely align.

One general implication of the above is that there may not be a unified theory of figurative language use and understanding, precisely because the reasons for using different tropes, and the mental processes involved in understanding metaphor, metonymy, irony, and so on are quite different and cannot be subsumed under a single umbrella that is distinct for figurative language alone.

B. Traditional Theories and Empirical Results

Following the traditional belief about differences between literal and figurative language, psycholinguists have focused a great deal on examining the possibility that figurative language is understood after some sort of preliminary analysis of an expression's literal meaning (Gibbs, 1994, 2002). The most famous, and now traditional, view of how listeners understand nonliteral meaning comes from H. Paul Grice's theory of conversational implicature (Grice, 1989), often dubbed the "standard pragmatic" view. Grice argued that the inferences needed to understand nonliteral meaning are derived from certain general principles or maxims of conversation that participants in talk – exchange are mutually expected to observe (Grice, 1989). Among these are the expectations that speakers are to be informative, truthful, relevant, and clear in what they say. When an utterance appears to violate any of these maxims, as do many of the figurative expressions in the opening newspaper article, listeners, or readers, are expected to subsequently derive an appropriate "conversational implicature" about what the speaker intended to communicate in context given the assumption that he or she is trying to be cooperative.

The results of many psycholinguistic experiments have shown the standard pragmatic view to be incorrect as a psychological theory (see Gibbs, 1994; Glucksberg, 2001). Numerous reading-time and phrase classification studies demonstrate that listeners/readers can often understand the figurative interpretations of metaphors, irony/sarcasm, idioms, proverbs, and indirect speech acts without having to first analyze and reject their literal meanings when these expressions are seen in realistic social contexts. For instance, people can read figurative utterances (i.e., "You're a fine friend" meaning "You're a bad friend") as quickly, sometimes even more quickly, as literal uses of the same expressions in different contexts, or equivalent nonfigurative expressions. These experimental findings demonstrate that the traditional view of figurative language as *always* requiring additional cognitive effort to be understood has little psychological validity.

But the idea that people can use context to infer figurative meaning without a literal analysis of an expression has been criticized on various grounds. First, there has been misunderstanding of the claim that figurative language can be understood "directly." This suggestion does not imply that people do not process the meanings, literal or otherwise, of the individual words in each expression. The work showing that people can process many instances of figurative language as quickly as they do nonfigurative speech only implies that a complete analysis of an expression need not be completed before any interpretation of its intended figurative meaning can begin (Gibbs, 2002).

Second, some studies have found evidence that people take longer to process figurative language than corresponding literal speech, exactly as would be predicted by the traditional view (Blank, 1988; Giora, 2002; Schwoebel, Dews, Winner, & Srinivas, 2000). Yet in at least some cases, the contexts used in these studies were relatively weak in supporting figurative meanings. For instance, remarks like "You're just in time" took longer to read in ironic context (i.e., when someone was quite late) than in literal ones (Giora, Fein, & Schwartz, 1998), especially when the irony was unexpected. But in other studies, the context in which an ironic remark appeared set up an ironic situation so that the speaker's utterance was easily understood as having ironic meaning and took no longer, and occasionally less time, to process than literal statements (Gibbs, 1986a, 1986b). Similar effects have been reported in regard to metaphor understanding where some contexts set up metaphorical conceptualizations of topics that make following metaphoric utterances easy to interpret (Gentner, Imai, & Boroditsky, 2002; Pfaff, Gibbs, & Johnson, 1997). People may still need to draw complex inferences when understanding some figurative statements, but part of these inferences can occur before one actually encounters a figurative utterance. Other studies show that familiar conventional instances of figurative language (e.g., "John kicked the bucket," "A fine friend you are") take less time to interpret than do novel figurative expressions (Giora et al., 1998; Temple & Honeck, 1997).

Listeners may take longer to understand a novel expression because of the difficulty in integrating the figurative meaning with the context and not because listeners are first analyzing and then rejecting the expression's literal meaning (Schraw, 1995; Shinjo & Myers, 1987). For these reasons, we simply should not infer that the literal meaning for an entire phrase or expression must have been analyzed simply because people take longer to read novel instances of figurative language than to process either familiar figurative expressions or equivalent literal statements (Brisard, Frisson, & Sandra, 2000). Bowdle and Gentner (2005) also caution that equating conventionality with directness of processing may be an oversimplification. The processing required to interpret novel figurative language depends on many factors, including grammatical form, context, and whether different instances are related to preexisting figurative schemes of thought (Bowdle & Gentner, 2005; Gibbs, 1994), enough so that even novel expressions may require as little time to understand as do conventional figurative utterances.

Another body of research has suggested that on-line studies may be better indicators of literal meaning activation than are more global measures of utterance comprehension, such as reading time and phrase classification techniques (Brisard et al., 2000). These on-line studies usually examine the activation of literal and figurative meanings at different points during and at the end of figurative utterance comprehension. For instance, one research project examined comprehension of familiar and less familiar metaphorical expressions (Blasko & Connine, 1993). Participants in these experiments heard different sentences and made lexical decisions at various times to visually presented word strings. For instance, as participants heard the sentence "The belief that hard work is a ladder is common to this generation," they were visually presented a letter string immediately after hearing the word "ladder." The letter string visually presented was related to some aspect of the sentence's literal meaning (e.g., "rungs"), a letter string related to the sentence's metaphoric meaning (e.g., "advance"), or a control word unrelated to the sentence (e.g., "pastry"). The results revealed that participants were equally fast in responding to the literal and metaphorical targets, which were both faster than the latencies to the controls. This was true both when participants made their lexical decisions immediately after hearing the critical word (e.g., "ladder"), and when the same decisions were made 300 ms after hearing the critical word. However, when participants made these same types of lexical decisions to literal and metaphorical targets having heard less familiar expressions, such as "The thought that a good professor is an oasis was clung to by the entire class," only literal targets were primed immediately after hearing the critical word (e.g., ‘‘oasis"), while responses to the metaphorical targets were facilitated only 750 ms after the critical word.

But these studies have one important methodological flaw in their equating different aspects of meaning (word vs. phrasal) with response times to literal (word) and metaphoric (phrasal) targets. For example, in Blasko and Connine (1993) the literal target "rung" is a simple semantic associate of the word "ladder," while the metaphoric target "advance" only relates to the general meaning of the entire expression. This makes it difficult to conclude anything about the time-course under which literal meanings of an entire sentence are activated compared to figurative meanings of these expressions. Even if one conceives of literal meaning as only relating to individual word meaning, this study does not compare activation of literal word meanings with figurative word meanings. Moreover, the words used as literal and metaphoric targets do not seem to reflect very distinctive literal and figurative meanings. The literal target "rung," for instance, is related to the idea of advancing (i.e., the figurative target) given that climbing ladders, even literally speaking, is an instance of advancing along some physical path. We believe these problems plague a good deal of the studies using lexical priming techniques to examine figurative language processing.

A different issue with many studies is the assumption that the activation of a particular meaning (i.e., literal or idiomatic) reflects the output of entirely different linguistic processes. The possibility remains that activation of different kinds of meaning (i.e., literal or idiomatic) reflects different types of meaning accessed by a single linguistic process. The fact that psycholinguists label one kind of meaning as "literal" and another "figurative" does not necessarily indicate that different processes operate (i.e., a literal processing mode and a idiomatic or figurative processing mode) to access these meanings (either in a serial or parallel manner). There are many types of figurative meaning, including metaphoric, idiomatic, metonymic, ironic, satirical, proverbial, hyperbolic, oxymoronic, and so on. Scholars often assume within the context of a single set of studies that there are two processes at work during figurative language understanding, such as literal vs. idiomatic, literal vs. metaphoric, or literal vs. ironic. Yet if there are numerous types of meaning, must there be dozens of types of linguistic processes all at work, or potentially at work, when language is understood? Psycholinguists have not addressed this question primarily because they focus too narrowly on only one kind of figurative meaning against a simple view of literal meaning.

A related type of study that examined the time-course for understanding literal and figurative interpretations of simple sentences used a signal, speed-accuracy trade-off procedure (McElree & Nordlie, 1999). Participants were presented strings of words, one at a time, at a rate that approximated fast reading (250 ms/word). The final word in each string produced a literal (e.g., "some tunnels are sewers"), a figurative (e.g., "some mouths are sewers"), or a nonsensical interpretation (e.g., "some cattle are sewers"). Participants judged whether each word string was meaningful when a tone appeared at varying times after the critical, last word. No differences were found in the comprehension speed for literal and figurative strings. McElree and Nordlie argued that the lack of time-course differences is inconsistent with the claim that figurative interpretations are computed after a literal meaning had been analyzed. In general, the time-course data presumably support the idea that literal and figurative interpretations are computed in parallel.

But we question whether the null results (e.g., no difference in processing literal and figurative sentences) obtained in these experiments necessarily provide evidence in favor of a parallel processing model. The activation of a particular meaning (i.e., literal or idiomatic) is assumed to reflect the output of entirely different linguistic processes. Once again, the possibility remains, however, that activation of different kinds of meaning (i.e., literal or idiomatic) may arise from a single linguistic process.

C. New Models and Findings

The continuing debates over the traditional view of figurative language understanding have led to the development of several alternative theories, specifically focused on the role of context in figurative language processing. These new models generally aim to describe the influence of context on figurative language processing at a more fine-grained level than the earlier proposals. Thus, the newer models suggest when and how context prompts figurative meanings during word-by-word linguistic processing. At the same time, these newer models attempt to offer general accounts that may apply to all aspects of figurative language, compared to most theories that aim to describe individual tropes (e.g., metaphor, irony, proverbs). Although these models recognize that some trope-specific types of processing may be necessary, they suggest that some obligatory linguistic processes operate with all types of figurative language.

Perhaps the most prominent of these new models is the "graded salience hypothesis" (Giora, 2002). This account specifically claims that context functions to constrain figurative meanings only after salient word or phrase meanings have already been accessed. Salient word or phrase meanings are not necessarily "literal" meanings. Instead, salient meanings reflect the most common, conventional use of a word or phrase. Unlike the standard pragmatic view, however, context may facilitate activation of figurative meanings before people analyze the semantic, or literal, meanings of the entire linguistic expression. For instance, processing familiar metaphors (e.g., "step on someone's toes") should activate both of their literal (e.g., foot) and metaphoric (e.g., offend) meanings, even when these metaphors are seen in appropriate discourse contexts. Processing unfamiliar metaphors (e.g., "Their bone density is not like ours") may, on the other hand, only initially activate their literal meanings, as these are most salient.

Different empirical studies, ranging from reading-time to word-fragment completion experiments, support this general idea for how people interpret different kinds of figurative language, in addition to how jokes may be understood. For example, consider the findings of a set of studies looking at irony comprehension (Giora & Fein, 1999). These studies examined people's understanding of familiar (e.g., "Very funny") and less familiar (e.g., "Thanks for your help") ironies in comparison to literal uses of the same expressions in appropriate contexts. Participants read stories ending with either literal or ironic remarks. After reading the final sentence, participants were presented with a letter string and had to quickly respond whether that string was a meaningful word. For instance, after reading the statement "Thanks for your help," participants were presented with either a ironic test word (e.g., "angry") or a literal test word (e.g., "useful"). These test words were presented either 150 or 1000 ms after participants read the final statements.

The results showed that when people read less familiar ironies they responded faster to the literal test words than to the ironic test words in the 150 ms condition, but there were no differences in the lexical decision times to the literal and ironic test words after 1000 ms. In contrast, the literal and ironic test words were responded to equally fast after both 150 and 1000 ms when people read familiar ironies. This pattern of data suggests that when people read familiar ironies both literal and ironic meanings are quickly accessed, but only literal meanings are initially activated when people read less familiar ironic statements. Although Giora and Fein (1999) favor a salience-first processing model, as opposed to the standard pragmatic account, their results support the idea that salient meanings, of perhaps both words and sentences, are always accessed first. In this way, the graded salience view is similar to modular views of linguistic processing in which context operates to narrow appropriate meaning after some initial context-independent word and phrase meanings have been activated.

One difficulty with the graded salience view is that is unclear what defines a word's, or expression's, salient meaning. Giora (2002) suggests, "The salient sense of a word, or an expression, is the one directly computable from the mental lexicon irrespective of inferences drawn on the basis of contextual information" (P. 18). Salience is a graded notion, and includes senses that are more frequent, conventional, or prototypical/stereotypical. The best empirical method for assessing the salient meaning of any word is to use standardized norms such as word frequency and word familiarity, although these alone do not necessarily indicate which of several alternative senses of a word are most salient. Ordinary speakers can, however, be asked to judge the frequency or familiarity of alternative word senses to obtain a measure of salience. Giora (2002) also suggests other behavioral tasks may be employed to assess salient meanings such as asking people to write down the meanings of words, or phrases, that "came to mind first" (p. 22), or to provide speeded responses to probes related and unrelated to words placed in neutral contexts. In general, however, it is not clear that these different methods all lead to the same salient meaning for individual words and phrases, even for a single person.

A different problem with the graded salience view is that it posits automatic activation of both salient word and phrase meanings. The motivation for this facet of the proposal comes from the fact that the conventional meanings of certain phrases, such as "kick the bucket" (meaning "to die"), are automatically activated even when the context specifies a different interpretation (e.g., a dairy farmer striking his foot against a pail).

Yet according to the graded salience hypothesis, the salient meanings of individual words should also be automatically activated regardless of context. Thus, the salient meaning of the word “kick” should be quickly accessed. But this salient word meaning differs from the putative salient meaning of the entire phrase (e.g., “to die”). It is unclear how this conflict is resolved or whether context comes into play to determine contextually appropriate word meanings before conventionalized phrasal meanings are accessed.

A related recent theory of figurative language processing claims that the language processor initially accesses an interpretation that is compatible with both a word’s literal and figurative meanings (Frisson & Pickering, 2001). Consider the verb “disarmed” in “Mrs. Graham is quite certain that they disarmed about every critic who was opposed to spending more money on art.” The “underspecification model” assumes, for example, that the initial meaning recovered when reading the verb “disarmed” in any context is underspecified as to whether it refers to removing literal or figurative arms. Over time, however, the language processor uses context to hone in on the word’s appropriate meaning, where the honing in process is faster when the preceding context is strong and slower when the preceding context is neutral.

Support for the underspecification model comes from several eye-movement studies. In one study, Frisson and Pickering (2001) examined people’s processing of ambiguous verbs, such as “disarmed” in the above sentence. The eye-movement data showed that the processing difficulty with the subordinate sense of “disarmed,” relative to when the word was used in a literal, dominant sense (e.g., “After the capture of the village, we disarmed about every rebel and sent them to prison”), did not emerge until *after* the critical verb wasmread. Thus, context reduces processing difficulty, but the difference did not emerge until much after the verb was seen. Frisson and Pickering suggest that people did not initially access either a specific sense or several senses for an ambiguous verb. Instead, readers initially recovered a general, underspecified meaning for the verb and then created a furtherconcrete instantiation of its meaning later on. According to the underspecification model, then, context does not operate to judge between different word meanings, but functions to change an underspecified, or highly general meaning, into a specific interpretation.

A different set of studies in support of underspecification investigated processing of sentences containing place-for-institution metonymies such as "That blasphemous woman had to answer to the convent" by measuring participants' eye-movements as they read (Frisson & Pickering, 1999). Results showed that people were as fast to understand these familiar metonymies as to read literal sentences, and that processing unfamiliar metaphors took more time than did reading compatible literal sentences. Thus, figurative language processing need not be delayed for familiar metonymies. A second study showed similar findings for familiar place-for-event metonymies such as "A lot of Americans protested during Vietnam," and unfamiliar ones such as "A lot of Americans protested during Finland." Frisson and Pickering argued that the overall findings do not support either a literal-first or figurative-first model, but fit best with a model where a single underspecified representation that is compatible with both literal and figurative (e.g., metonymic) senses. Eventually, context comes in to hone the very general interpretation into a contextually appropriate meaning.

The underspecification model does not assume that different linguistic processes must exist for different meaning products (i.e., literal vs. figurative uses of words) to arise during on-line linguistic understanding. In this way, the putative distinction between literal and figurative senses of a word is irrelevant, at least in terms of ordinary processing. However, similar to the graded salience model, the underspecification model embraces a modular view of linguistic processing, at least in the sense that lexical access is encapsulated from contextual effects. But similar to the graded salience view, the underspecification model suffers from the problem of not being able to specify what constitutes the initial, underspecified meaning that is accessed when a word is first encountered. Many linguists reject the underspecification view precisely because they have failed to discover senses that are rich enough to capture the wide range of meanings (up to 100 for some polysemous words) many words possess (Gibbs, 1994). More generally, both the graded salience and underspecification views face the challenge of demonstrating consistent bottom-up activation of context-free word meanings even in the presence of strong supporting context.

Finally, a different model of figurative language understanding embraces the notion of constraint satisfaction, an idea that has gained much support in psycholinguistics and cognitive science (Katz & Ferratti, 2001;

Katz, 2005). When people comprehend a text, or a figurative utterance, they must construct an interpretation that fits the available information (including context) better than alternative interpretations. The best interpretation is one that offers the most coherent account of what people are communicating, which includes meanings that best fits with certain other information and excludes meanings that do not fit this other information. Under this view, understanding a figurative utterance requires people to consider different linguistic and nonlinguistic information that best fits together to make sense of what a speaker or writer is saying. Constraint satisfaction models are computationally efficient, and perhaps psychologically plausible, ways of showing how different information is considered and integrated in everyday cognition.

Katz and Ferretti (2001) argue that a "constraint satisfaction model" provides the best explanation for experimental data on proverb understanding. They employed a self-paced moving window paradigm to show that context affects people's immediate reading of familiar (e.g., "Lightning never strikes the same place twice") and unfamiliar proverbs (e.g., "Straight trees have crooked roots") that have both well-formed literal and figurative meanings. Familiar proverbs were understood more easily than unfamiliar expressions, and the speed-up in processing for familiar proverbs occurred as soon as the second word of the expression was read. But the first words of unfamiliar proverbs were read more quickly in contexts supporting their figurative, rather than literal, meanings. Yet the analysis of an unfamiliar proverb's figurative meaning was not always complete when the last word was read.

These findings support a constraint satisfaction model by positing how different sources of information (i.e., syntactic, lexical, conceptual) compete for activation over time in parallel. Constraints interact to provide probabilistic evidence in support of various alternatives with the competition ending when one alternative fits best. For example, when reading an unfamiliar proverb, people immediately focus on a literal interpretation because there is less competition from other sources of information supporting a figurative meaning. Similarly, familiar proverbs are easier to process than unfamiliar expressions because there is more information available from the context and the words in familiar proverbs to support a figurative interpretation.

Another test of the constraint satisfaction view examined people's immediate understanding of expressions like "Children are precious gems" as having metaphoric (children are valuable) or ironic (children are burdens) meaning (Pexman, Ferretti, & Katz, 2000). Several sources of information could induce either the metaphoric or ironic meaning, including the occupation of the speaker, whether the statement was counterfactual to information in the previous discourse, and the familiarity of the expression. Results from an on-line reading task (i.e., moving window) demonstrated that the "A is B" statements were initially read as metaphors, but that the speaker's occupation and counterfactuality of the statement given the previous context play an early role in processing, thus slowing processing at the space following the statement or by the time the first word of the next statement is read. Furthermore, knowing that a speaker is often associated with irony slows down reading of the first word in the following statement if the context leads one to expect a metaphoric reading, yet acts immediately to speed up processing right after the target statement if the context induces an ironic meaning. The complex interaction between the three sources of information is consistent with the idea that understanding whether an expression is meant metaphorically or ironically depends, similar to other aspects of language, on multiple sources of information being examined and interpreted continuously during on-line reading (McRae, Spivey Knowlton, & Tannenhaus, 1998).

Related findings using a moving window paradigm showed that context modulated relative processing of literal and ironic statements (Ivanko & Pexman, 2003). When context induced neither a literal or ironic bias, reading times for literal and ironic utterances were roughly equivalent, with faster reading times more locally for the fifth word of the target statements. When the context led to a bias for literal criticism, ironic remarks were read more slowly than literal ones. Once again, there are complex interactions between the type of context and the speed with which figurative utterances are understood, such that literal readings of utterances, or salient ones, are not obligatory in all cases. This pattern is most consistent with probabilistic, constrain-satisfaction models of figurative use, and is inconsistent with modular approaches to linguistic processing.

D. Indeterminacy of Figurative Meaning And Processing

The important emphasis on on-line-processing figurative language in experimental psycholinguistics often ignores exactly what people have understood when they seem to have successfully comprehended a particular figurative expression. For the most part, psycholinguists and others tacitly assume that any figurative statement can be paraphrased by a linguistic expression that states in literal terms what people must have attempted to communicate when speaking figuratively (e.g., “blow your stack” means “to get very angry”). The reduction of figurative meaning to simple, short linguistic paraphrases in psycholinguistics is reasonable in the context of designing experimental studies that, for instance, contrast figurative language processing with nonfigurative, or sometimes literal, understanding.

However, the belief that figurative expressions can be readily paraphrased misconstrues the complexities of what many figurative expressions actually communicate in real-life contexts. Studies show, for example, that when people read “John blew his stack,” they readily infer information about the cause, intentionality, and manner by which John got angry, inferences they did not draw when they read literal paraphrases of similar length such as “John got very angry” (Gibbs, 1992). Furthermore, reading idioms in contexts that violate any of these inferences slows down processing for these phrases, but not so for literal paraphrases (Gibbs, Bogdonovich, Sykes, & Barr, 1997). These empirical findings strongly suggest that even highly conventional metaphors, which are often incorrectly assumed to be dead metaphors or long lexical items, convey rich conceptual and pragmatic information, more so than do so-called literal paraphrases.

This conclusion about the richness of figurative meaning comes as no surprise to many interdisciplinary metaphor scholars who have long argued that metaphor, and many other tropes, are “pregnant with meaning.” In fact, a large body of research has discovered that different forms of figurative language communicate a wide variety of propositional, social, and affective meanings, or pragmatic effects. Verbal irony, for instance (e.g., saying “This is fantastic” when losing one’s keys), has been shown to both enhance and diminish the condemnation expressed by an individual relative to speaking more directly (Colston, 1997). By saying something positive about a negative situation, the situation is made to look worse relative to saying something directly negative (e.g., “This is just awful”), which enhances the speaker’s condemnation. Verbal irony, along with hyperbole (e.g., “He was so hungry he ate the table with his meal"), and understatement (e.g., "This might require a bit of work" about a huge task), also express predictably variable degrees of humor, expectation/reality deviance demonstration and speaker protection (Colston & O'Brien, 2000a, 2000b; Colston, 2002). Hyperbole expresses surprise either through increasing hearers' attention toward expectation/reality discrepancies via the distinctiveness of the inflation, or by an audacity demonstration process whereby a speaker breaks with conversational congruity to make a point (Colston, in press; Colston & Keller, 1998).

One form of ironic discourse, called rebuttal or ironic analogy, performs the dual pragmatic functions of argument and social attack (Whaley & Holloway, 1996; Colston & Gibbs, 1998; Colston,1999, 2000a). So if a speaker says, for instance, "calling Chili's just another restaurant is like saying the Great Wall of China is just a fence," she causes the hearer to map the ironic structure of the base, "saying the Great Wall of China is just a fence", onto the target, "calling Chili's just another restaurant." This acts to argue against the proposition in the target, as well as to belittle the proponent(s) of that proposition.

Many forms of figurative language also bolster persuasiveness and the social standing of speakers (Holtgraves, 2001; Sopory & Dillard, 2002). A "truth externalization" process is particularly well performed by proverbs, metaphors and some idioms, for instance. Thus, a speaker who claims, "It is best to let sleeping dogs lie" relies on a cultural norm as expressed in the proverb to convey the best course of action in a potentially difficult situation (Gibbs, 2001; Curco, 2005). By using language that leverages a significant degree of meaning outside of a proposition directly proclaimed by a speaker, the speaker places the "truth" of the intended message outside of him/herself. This lending of objectivity can make the meaning seem stronger. A speaker's demonstrated skill in sheparding the intended message of a figurative utterance can also increase others' admiration, which can in turn additionally enhance the message. These and other similar processes can contribute to a more general "mastery demonstration" function where a figurative speaker can gain in their social standing by using figurative language (Gibbs & Izett, 2005). Indeed, many people have a positive subjective experience when they comprehend figurative language (for a review, see Colston, in press). Such a positive feeling can reflect well on a speaker and lead to many sociocognitive and persuasive effects (e.g., liking the speaker more, paying greater attention to what the speaker says subsequently, more strongly adopting the speaker's viewpoint, etc.) that can cascade and contribute to other of the effects discussed here.

Finally, research shows that some forms of figurative language evoke different kinds of emotional reactions. Thus, hearing ironic statements leads listeners to feel more intense emotions than when literal speech is heard (Leggitt & Gibbs, 2000). Sarcasm, rhetorical questions, and overstatements all evoke relatively negative emotions, compared to understatement and satire. People also tend to speak metaphorically more so when feeling intense emotions, something that listeners readily pick up on in many conversations and attribute affective meanings to speakers' messages (Gibbs, Leggitt, & Turner, 2003). One large study demonstrated, more generally, that different types of figurative language can fulfill as many as 20 different discourse goals, including many of the social and affective effects described here (Roberts & Kreuz, 1994).

Not surprisingly, inferring pragmatic effects may come at a processing cost. But what determines the stopping point for the various indeterminate aspects of figurative meanings? One well-known proposal from linguistic pragmatics suggests that there is a trade-off between the amount of cognitive effort put into linguistic understanding and the cognitive effects, or meanings, that are inferred (Sperber & Wilson, 1995), a theory that applies to all aspects of linguistic communication, not just figurative language processing. Relevance theory generally claims, again, that interpretation of figurative language occurs in the same way as with any other nonfigurative utterance. A listener stops processing when he thinks that every further implication he could get is not worth the effort it takes to obtain these additional cognitive effects. Sperber and Wilson (1995) claim that newly presented information is relevant in a context only when it achieves cognitive effects in that context, and the greater the cognitive effects, the greater the relevance. They specifically define a notion of "optimal relevance" that outlines what listeners look for in terms of cognitive effort and effect: an utterance, on a given interpretation, is optimally relevant if and only if (a) it achieves enough effects to be worth the listener's attention, and (b) it puts the listeners to no gratuitous effort in achieving these effects.

Consider the metaphorical utterance "My surgeon is a butcher." Listeners generally have immediate access to stereotypical knowledge about surgeons and would normally infer that the speaker here means, "My surgeon is crude and sloppy in his practice." Speaking loosely like this requires that speakers have in mind some further idea or cognitive effect beyond the single thought "My surgeon is crude and sloppy in his practice." For instance, the speaker may wish to convey an image of surgeons that is beyond most people's experience and will expect the listener to make some effort toward exploring a wide range of cognitive effects (e.g., having to do with the nature of surgeons, their imprecision, their insensitivity toward dealing with human beings, and perhaps their appearance and demeanor). These implications are relatively weak, but they best resemble the speaker's thought about his surgeon. Understanding this range of weak implications may require additional cognitive effort on the part of the listener, but this is offset, according to the principle of relevance, by extra effects not achievable by saying directly "My surgeon is crude and sloppy in his practice." In general, metaphorical utterances, like all figurative and indirect language, are simply one means of optimizing relevance in verbal communication.

Very few psycholinguistic studies have explicitly explored the trade-off between cognitive effort and effects in figurative language processing. One study suggests that there must be extra processing associated with understanding a well-chosen metaphor (Noveck, Bianco, & Castry, 2001). Yet it is not clear how to operationalize the idea of individual metaphorical, or figurative, meanings within the "more cognitive processing more cognitive effects" hypothesis. Consider the stock metaphor "Some jobs are jails." There are a variety of meanings that people may understand when reading this expression, including that some jobs are poorly paid, confining, stifling, unpleasant, demoralizing and so on. But how does one actually distinguish between these impressions to clearly establish which meanings are independent in order to test the idea that more cognitive processing equals more cognitive effects? This problem is complicated by the possibility, as noted above, that listeners may draw a range of pragmatic effects, or weak implicatures, from figurative utterances. There are also cases where people can put a good deal of cognitive effort into understanding a speaker's utterance without gaining appropriate cognitive effects. For example, where people assume that the producer of a metaphor was a famous poet, they put in a good deal more effort to try to understand anomalous phrases, such as "A scalpel is like a horseshoe," than when told that these phrases were randomly generated by a computer program (Gibbs, Kushner, & Mills, 1990). Finally, people may also infer complex figurative meanings with little cognitive effort, or at least less time than is needed to comprehend corresponding nonfigurative expressions, as shown by Gibbs (1992).

In general, it is impossible to predict the processing effort needed to comprehend figurative utterances given the number, or types of cognitive effects than may arise from interpreting these statements. It may be the case, as Noveck et al. (2001) argue, that some figurative expressions, such as certain novel metaphors, may take longer to process than synonymous nonfigurative expressions, if these are encountered in neutral contexts, precisely because of the additional cognitive effects they communicate. But proving this point will require an independent measure of the cognitive effects that utterances convey. We see this as one of the great challenges for figurative language scholars, as well as for all psycholinguists.

E. Examining the Cognitive Effort and Effects Trade-Off

Despite some of the difficult questions regarding the nature of cognitive effects, we believe that the time is ripe for psycholinguistic research on the trade-off between effort and effects during figurative language processing. Our claim here is that present debates over whether figurative language is understood directly or indirectly should evolve into a more systematic examination of the complex interactions between many cognitive and linguistic factors associated with any psycholinguistic act. One way to begin this type pf exploration is to adopt an old tetrahedral model of cognitive processes (Jenkins, 1979), which suggests that several factors must shape processing, including (1) the participants (e.g., their abilities, interests, beliefs, motivations, goals), (2) the understanding task (e.g., understanding to solve a problem, make a decision, remember something, be emotionally affected by something said), (3) the criterial task (e.g., different measures of cognitive processes and product), and (4) the materials (e.g., type of language, modality of presentation). Fortunately, there is a fair amount of research relevant to each of these factors in regard to figurative language use, even if at present these findings have not been placed within a larger theoretical framework.

1. Participants

There are a variety of participant variables that can influence processing fluency for figurative language. For example, if a speaker is known to be the type of person who regularly uses verbal irony, based on occupation or gender for instance, and if the situation has been set up to likely create ironically intended utterances, then ironic utterances will be comprehended relatively fluently, as several studies have demonstrated (Katz & Pexman, 1997; Pexman & Olineck, 2002; Katz, Piasecka, & Toplack, 2001). Gender is another important variable, given research showing that men tend to use figurative language in describing other people's emotions, while women use figurative speech more in talking about their own feelings (e.g., "I would feel like my heart would just jump out of my chest … . ") (Link & Kreuz, 2005). The relationship between speakers (e.g., close friends vs. strangers), their social status, occupation, geographic origin, religious or political background, ethnicity, and personalities, show effects on comprehension of a diverse range of figurative forms, including metaphor, irony, metonymy, proverbs, idioms, indirect requests, analogies, litotes, and metaphorical gestures (Colston & Katz, 2005). To note just a couple of examples, consider the normal ironic banter that often accompanies groups of friends (Gibbs, 2000; Pexman & Zvaigzne, 2004) or the quintessential British form of understatement (e.g., "He clearly has issues" in reference to a suicidal character).

Another emerging characteristic of participants that has been shown to influence figurative language processing is their past and current embodied experiences (Gibbs, 2005). For example, research shows that people's previous bodily experiences of hunger partly predicts their use and understanding of metaphorical expressions about difference forms of desire, as seen in statements like "I hunger for fame" or "I craved her affection" (Gibbs, Lima, & Francuzo, 2004). In another series of studies on metaphorical talk about time, students waiting in line at a café were given the statement "Next Wednesday's meeting has been moved forward two days" and then asked "What day is the meeting that has been rescheduled?" (Borodistky & Ramscar, 2002). Students who were farther along in the line (i.e., who had thus very recently experienced more forward spatial motion) were more likely to say that the meeting had been moved to Friday. Similarly, people riding a train were presented the same ambiguous statement and question about the rescheduled meeting. Passengers who were at the end of their journeys reported that the meeting was moved to Friday significantly more than did people in the middle of their journeys. Although both groups of passengers were experiencing the same physical experience of sitting in a moving train, they thought differently about their journey and consequently responded differently to the rescheduled meeting question. These results suggest how ongoing sensorimotor experience has an influence on people's comprehension of metaphorical statements about time.

2. Orienting Task

In many communicative settings, people are not given, nor are they following any explicit directions or rules that might orient them to process or comprehend what is said or written in a particular way. Exceptions to this would be in those occasional situations where orienting rules or directions have been given or are being followed, (e.g., a law clerk is told to scan through court transcripts looking for when a defendant said...). People would, though, likely adhere to implicit rules that might affect figurative language comprehension and processing (c.f., eavesdropping vs. listening to poetry).

For example, asking people to solve problems with metaphoric language, simply understand metaphors, recall metaphors, or produce metaphor may all produce varying empirical findings in regard to the relative primacy of metaphor to nonmetaphoric language. Thus, research shows that people's decisions about common dilemmas are strongly shaped by the presence or absence of metaphor (Robins & Mayer, 2000). When a metaphor is critical to frame, or understand, a problem, people readily use that information in making decisions about a common dilemma. But when metaphor is not necessary to understanding a dilemma, the presence of such language adds ambiguity to people's decision-making process. On the other hand, studies also show that the persuasiveness of metaphor depends more on the way such language resonates with a person's own preferences than it does with whether the metaphor is needed to frame a topic (Ottati, Rhoads, & Graesser, 1999).

In a different context, although people clearly understand familiar metaphors faster than they do unfamiliar ones, they recall these two types equally well (Blasko & Brihl, 1997). Finally, asking people to verbally describe the conceptual connection between two word often yields metaphoric descriptions, which take longer to produce than do nonmetaphoric descriptions

(Flor & Hadar, 2005), a result that is contrary to the typical pattern in reading time studies showing that metaphors do not generally take more time to process than nonmetaphorical statements. Once again, the orienting task given to experimental participants can have an important effect on whether figurative language is seen as primary compared to nonfigurative speech.

At the same time, orienting tasks can powerfully adjust the fulcrum location upon which language comprehension tips toward the more figurative or less figurative. For instance, when operating under the criterion of achieving genuine or deep understanding, listeners/readers can use figurative comprehension to fertilize rich interpretation. Many heady, moving experiences of language comprehension are evidence of this (e.g., hearing powerful speeches, emotional song lyrics or poetry, highly apt metaphors or other figures). Conversely, when listening/reading for less cooperative and indeed, combative purposes (e.g., as in arguments, debates), people will often scramble for the golden fleece of a "literal meaning," to serve those purposes (e.g., for preparing rebuttals, to find weaknesses or attack points in others' comments). One study for instance found evidence for such a link between criterial task and degree of "literal mindedness." When people were placed in high-stress situations, as are often the case in arguments, disagreements, debates, etc., their ability to comprehend figurative language subtly broke down (Colston, 2005b). Barr and Keysar (2005) also argue that people tend to interpret figurative expressions egocentrically, and thus do not take common ground information immediately into account, when they are under time pressure.

3. Criterial Task

The best method for assessing comprehension or interpretation of figurative language has always been a significant source of concern in psycholinguistics. Tasks that have used off-line measures (e.g., rating studies, judgments of metaphor aptness, memory tasks) as indicators of figurative language comprehension have often been criticized for their inability to distinguish processes that might take place during reading or hearing in the comprehension process vs. those that might occur later in the processing stream. Reading time measures were long considered superior because they could use overall reading time as a more precise, and presumably outside of subjective control, indicator of on-line processing – relying upon the assumption that, all else being equal, longer reading times indicated greater processing. But reading time studies also differ in their specific task requirements. Some experiments ask participants to simply read individual sentences in a story, and push a button as soon as the expression on the computer screen has been understood. Yet other studies ask people to sometimes read an expression, such as a figurative remark, and make a speeded judgment as to whether it fits within the preceding story context. As it turns out, judgments of appropriateness or relatedness often result in longer comprehension times for figurative expressions compared to literal ones (Schwobel et al., 2001; Temple & Honeck, 1997). But figurative and literal expressions can be read equally fast when only simple comprehension time is measured (Gibbs, 2002). Thus, the precise task used leads to different results with very different theoretical implications.

More recently, more sophisticated mechanisms have enabled word-to-word reading time measures with moving scanning windows that readers control, and eye-tracking measures that remove unnatural reader responses altogether from the reading/measurement. This progression in research methodologies has been viewed as an improvement in our ability to tap into figurative language processing, and undoubtedly it has afforded greater precision. But often overlooked in this perspective is the potentially remaining disconnect between even the very precise eye-movement measures and what one genuinely and subjectively experiences as comprehension, as if that is ever even a delineated, all-or-none accomplishment, universal across all kinds of language, interlocutor types, goal-situations, etc. (but see Rayner and Pollatsek, this volume for a different perspective). Eye movements, although rich in their potential, may not necessarily be deterministically related to states of comprehension in completely reliable ways (e.g., a reader may pause and stare at a random word while processing some text that is largely irrelevant to that word). Moreover, the problem that text comprehension, spoken language comprehension from say anonymous audio recordings, and genuine conversational comprehension with known interlocutors are very different things is mostly overlooked given the primary emphasis on appropriate dependent measures. One possible solution would be to further increase the sophistication of "comprehension" measures, such as combining emotional response indicators, eyemovement trackers, video facial expression recordings and other measures in a linked time course measurement (Colston, 2005a).

Within interdisciplinary discussions of figurative language, scholars from fields outside of psycholinguistics often see psychological experiments as being rather distant from their own concerns with the deep, meaningful, complex interpretation of different forms of figurative discourse (e.g., metaphor, irony, metonymy). Psycholinguists' primary interest in immediate, fast, mostly unconscious mental processing of figurative speech ignores slower, more reflective aspects of linguistic interpretation. Gibbs (1994) suggested that figurative language understanding does not constitute a single event, or moment in time, but can exist along a continuum of temporal processing ranging from fast comprehension, slower interpretation, nonobligatory recognition (i.e., that statement I just heard was ironic) to reflective appreciation. In the past, scholars have mistakenly made theoretical claims about fast processing from slower, and consciously held, interpretations and appreciations, while psycholinguists, again, have mostly neglected the rigorous study of cognitive effects, or the products of figurative language understanding.

Yet once more, paying systematic attention to cognitive effects, both those that arise immediately during fast comprehension, and those that emerge more slowly during reading (and re-reading!) is critical to creating more comprehensive theories of figurative language use in different real world, communicative contexts.

4. Materials

There are several aspects of the materials that have been shown to have a strong impact on figurative language use and processing. First, as described above, the conventionality of a figurative expression plays a major role in the way it is processed. But conventionality is not a single dimension, given that several factors contribute to the impression that some utterance is conventional or novel, including its grammatical form, frequency in thelanguage, appropriateness to the specific context, and appropriateness for the speaker. For example, consider verbal irony. Conventional ironies will often take the form of rhetorical or tag questions (e.g., "wasn't that brilliant?", "that was brilliant wasn't it?"). As mentioned earlier, conventional ironies also often contain noun or verb modifiers, usually intensifiers (e.g., simply, utterly, just, absolutely, etc.). With regard to semantics, a highly conventional ironic pattern is to use utterances with positive meanings, usually to comment about negative situations. Other semantic conventions in verbal irony are to express agreement (e.g., "yup," "uh huh," "sure," etc.), to understate (e.g., "it seems to be snowing" said during a blizzard), to exaggerate (e.g., "her two foot tall husband," said about a short man) or to express the semantic content that was predicted or that might be expected in the situation (e.g., "soccer is an 'easy' game"), often when the situation has not gone as expected (e.g., the game is difficult). For example, the ironic phrase, "wise guy," is highly conventional in some American-English speaking communities because it is almost never used directly to state that a person is intelligent. Other utterances might have both ironic and literal conventional meanings. Examples here, again in some American-English speaking communities, are "oh, sure," "I'm sure," etc. Lastly, still other kinds of utterances are not at all conventionally ironic but might be used ironically in a given particular conversation.

The prototypicality of the material is also an important factor. The general pattern of results is that many kinds of processing (e.g., recognizing, recalling, reading, comprehending, etc.) are more readily accomplished to the degree that the target material they work upon is more prototypical. So, for instance, prototypical items of furniture (e.g., chair) are more easily and quickly recalled than less prototypical ones (e.g., lamp). More prototypical forms of syntactic structure (e.g., active) are more easily read than less prototypical ones (e.g., passive), etc. For verbal irony then, one can readily predict that more prototypical forms (e.g., positive semantic content, intensifying modifier, tag question) will be more fluently processed than less prototypical ones (e.g., negative semantic content, etc.). Indeed, the particular finding in this example has been born out by research (Gibbs, 1986a, 1986b, Kruez & Glucksberg, 1989; Utsumi, 2000).

To demonstrate this prototypicality influence, recall the earlier discussion on semantic conventions for verbal irony. Often ironic utterances will contain positive semantic content, but they will be used to comment about a negative situation (e.g., saying or writing, "Excellent, just what I needed," to complain about an unexpected extra workload). This frequently observed pattern of verbal irony illustrates a semantic contrast, which is a major characteristic, and indeed necessary condition, for verbal irony (Colston, 2000b). Ironic utterances that present strong contrasts between expectations and reality are funnier, more criticizing, less self protective, more expressive of surprise, etc., than verbal ironies that provide relatively weak contrasts (Colston & Keller, 1998; Colston & O'Brien 2000a, 2000b; Gerrig & Goldvarg, 2000). These pragmatic functions are not direct measures of comprehension fluency but they do indicate the power of the contrast variable on the expressiveness of the irony and strongly suggest a processing fluency difference between strong and weak contrasts. One might thus predict that, as usual with all else being held constant, the greater the contrast between expectations and reality, the easier verbal irony processing might be.

One major influence on processing fluency that is idiosyncratic to verbal irony is the clarity with which a mention, echo, reminder, or allusion to prior predictions or expectations that have been violated by occurring events can be achieved (Sperber, 1984; Kreuz & Glucksberg, 1989; Kumon-Nakamura, Glucksberg, & Brown, 1995). Indeed such a contrast between the semantic content of the utterance and its referent situation is in fact a hallmark of the ironic figurative form (Colston, 2000).

Metaphoric language also builds off from previous discourse structures that can quickly lead readers to metaphoric, as opposed to literal interpretations (Keysar, 1994), and in some instances are specific to this form of speech. Thus, contexts that describetopics in metaphorical ways make it easier to infer subsequent metaphoric utterances when the underlying conceptual metaphors are similar (Albritton, McKoon, & Gerrig, 1995), and more difficult to process when a new metaphorical utterance is based on a different conceptual metaphor (Langston, 2002). Although the vast amount of work on figurative language comprehension examines interpretation of a single utterance after a nonfigurative context, it is evident that different figurative contexts, and previously spoken figurative utterances, have a strong effect of on-line figurative language comprehension. This is one topic that demands further attention.

Finally, speakers use a wide variety of metalinguistic devices to indicate figurative intent. The phrase "strictly speaking" often accompanies metaphorical expressions (Goddard, 2004), and studies show that the presence of markers like "proverbially speaking" facilitates people's comprehension of proverbs (Katz & Ferratti, 2003). One direct way to affect the comprehension fluency of verbal irony is by providing or omitting markers for the irony

(Bryant & Fox Tree, in press). These markers can be controlled by a speaker or supplied in written context. For instance, nonverbal indicators of irony (e.g., gestures, facial expressions, etc.) can be performed or described (e.g., "she rolled her eyes and said"). Intonational patterns (e.g., nasal pronunciation, elongated phonemes, exaggerated pitch magnitudes, etc.), can be used or mentioned (e.g., "in a mocking tone he said"), etc. (Kreuz & Roberts, 1995). Speakers can also make simultaneous contradictory expressions to achieve expressional irony. For example a speaker could say, "Oh, yeah, he is brilliant," while making a gesture of tracing a circle around their ear, as if expressing that someone is crazy, and achieve an ironic perlocution. These factors would also contribute to the fluency of utterance processing, again to the extent that the speaker successfully uses them, that they are readily incorporated by the hearer, etc. In general, one might expect that the presence of such markers in the verbal and written situations would aid processing relative to their absence, again assuming equivalence of other influences.

5. Summary

Determining the constraints that shape the trade-off between cognitive effort and effects during figurative language understanding requires, in our view, the empirical study of how these various factors have their individual effects, and very likely interact in complex, even nonlinear, ways. The goals and motivations of individual speakers and listeners surely shape the cognitive effort expended when producing and understanding figurative language, and the specific orienting perspective, the methods used for assessing understanding, and the types of materials employed clearly determine the cognitive effects drawn in linguistic communication. Our hope is that specific recognition of thes factors will enhance attempts to create a more complete picture of the trade-off between cognitive effort and effects in figurative language use.

CONCLUSION

The complexities of figurative language processing are such that there may not be a single theory or model that explains how all aspects of figurative language are understood. Part of the reason for this conclusion is that figurative language does not constitute a homogenous kind of language that is necessarily used and understood in completely distinct ways from nonfigurative, or what some call “literal” speech. Of course, one message of this chapter is that it makes little sense to suggest theories of figurative language understanding, as different from “normal” discourse comprehension, unless there is a well-developed, and consistently applied, theory of literal language and meaning. Given the long history to provide a theory of literal meaning (Gibbs, 1994), and the failure to come up with a unified account of this kind of language, we frankly are doubtful whether any such proposal will come forward that is widely embraced by psychologists, linguists, and philosophers.

None of this implies that different aspects of figurative language have no special features, both in terms of the cognitive processes involved (e.g., cross-domain mappings for metaphor, determining the source of echos for irony, inferring part to whole relationships with metonymy) and the meaning products that arise from interpretive processes. We have argued that the study of both cognitive processes and effects, or products, is critical to future theoretical work on figurative language, and that exploring the real-time trade-off between effort and effects is one specific direction for new experimental studies. In this way, figurative language research should provide another arena within psycholinguistics more generally where the traditions of language as product and language as action perspectives may be bridged.

[This page is intentionally left Blank]

CHAPTER 13
EYE MOVEMENTS AND SPOKEN LANGUAGE COMPREHENSION

INTRODUCTION

Many everyday tasks require people to rapidly interrogate their visual surroundings. Reading a magazine, looking for a friend at a party, and making breakfast, all require people to frequently shift their attention to task-relevant regions of the visual world. These shifts of attention are accompanied by shifts in gaze, accomplished by ballistic eye movements known as saccades, which bring the attended region into the central area of the fovea, where visual acuity is greatest. The pattern and timing of saccades, and the resulting fixations, are one of the most widely used response measures in the brain and cognitive sciences, providing important insights into the functional and neural mechanisms underlying attention, perception, and memory. Eye movements are now an important measure in the perception–action literature, especially for studies examining allocation of attention in natural every day.

More recently, eye movements have become a widely used response measure for studying spoken language processing, in situations where participants comprehend and generate utterances that are about a circumscribed "visual world". The visual world paradigm has opened up relatively uncharted territory in language comprehension, including real-time sentence processing in children; the role of common ground in on-line processing; how listeners make use of disfluencies in real-time language processing; and how participants in a conversation coordinate their referential domains. Finally, the visual world approach has spawned a new family of studies investigating the interface between action and language and between vision and language.

The goal in this paper is to provide an introduction and overview to the rapidly growing literature on eye movements and spoken language processing, focusing on applications to spoken language comprehension. Major topics include the logic linking eye movements to spoken language processing, how eye-movement data are collected and analyzed, sample applications illustrating some of the paradigms, including comparisons to eye-movement reading studies, and associated experimental logics, and finally, concerns and limitations that arise in examining language in a circumscribed visual world.

A. Eye Movements

Eye movements are a behavior that can be measured and their measurement provides a sensitive means of learning about cognitive and visual processing. Although eye movements have been examined for some time, it has only been in the last few decades that their measurement has led to important discoveries about psychological processes that occur during such tasks as reading, visual search, and scene perception. Eye movements refer to the voluntary or involuntary movement of the eyes, helping in acquiring, fixating and tracking visual stimuli. Specific systems are used in maintaining fixation, when reading and in music reading. A special type of eye movement, rapid eye movement, occurs during REM sleep.

The eyes are the visual organs of the human body, and move using a system of six muscles. The retina, a specialized type of tissue containing photoreceptors, senses light. These specialized cells convert light into electrochemical signals. These signals travel along the optic nerve fibers to the brain, where they are interpreted as vision in the visual cortex.
Primates and many other vertebrates use three types of voluntary eye movement to track objects of interest: smooth pursuit, vergence shifts and saccades. These movements appear to be initiated by a small cortical region in the brain's frontal lobe. This is corroborated by removal of the frontal lobe. In this case, the reflexes (such as reflex shifting the eyes to a moving light) are intact, though the voluntary control is obliterated.

B. Spoken Language Comprehension

Curiously the study of speaking in its own right has been a relatively recent addition to the range of linguistics and applied linguistics specialisms. This is in spite of the popular assumption that knowing or learning a language centrally involves being able to 'speak' it, or the common claims that language pedagogy and linguistics prioritized the study and teaching of the vernacular. For most of the 20th century, speech was seen by linguists as only partially accessible to study: through phonetics and phonology; by studying idealized underlying competences, on the assumption that speech was transient and subject to the contingent influences of processing limitations of little linguistic interest; and thirdly through dialectology. Somewhat in parallel, although the approaches to language teaching developed since the Reform movement in the late 19th century have consistently made claims about the centrality of speech within their pedagogies, in fact speech was mainly of interest either because the oral medium is peculiarly appropriate to encouraging the unmediated and rapid processing of form-meaning pairings in a second language, along with conditions likely to favor memorization, notably opportunities for immediate feedback and frequent repetition; or else because one purpose for learning a second language that couldn't be ignored was to manage basic face-to-face service encounters while in the foreign country. However, perhaps because only towards the last quarter of the 20th century the particular patterns of speech and the nature of on-line processing became technically amenable to study, the actual forms of language studied and taught orally in second language classrooms were largely based on artefacts of the written language (such as sentence patterns, and scripted dialogues). Hence in spite of the fact that language teaching and linguistics have long claimed to place oral language at the center of its curricula, speech has been largely seen as similar to writing, but for the fact that it is processed orally.

Language comprehension is an important aspect of day to day functioning in adulthood. Comprehension of written and spoken language relies on the ability to correctly process word and phrase meanings, sentence grammar, and discourse or text structure. Difficulties in any of these domains can produce comprehension problems. Age-related memory declines have been reported in many studies comparing younger and older adults on language comprehension tasks. Therefore, it is believed memory capacity limitations in older adults may cause language comprehension problems (Wingfield and Stine-Morrow). In particular, age-related declines in the capacity of working memory to temporarily store linguistic information may be responsible for older adults' language comprehension problems. Older adults have typically been found to have smaller working memory spans than young adults and such span measures have been found to correlate with language comprehension measures. Van der Linden, and colleagues (1999) tested young and older adults on their ability to understand texts and recall sentences and words. They were also given a large battery of tests designed to measure processing speed, working memory capacity, and the ability to inhibit distracting thoughts. The analysis indicated that these three general factors (speed, working memory, inhibition) did account for age-differences in performance on the language processing tasks. Further, Van der Linden, and others concluded that "age-related differences in language, memory and comprehension were explained by a reduction of the capacity of working memory, which was itself influenced by reduction of speed, [and] increasing sensitivity to interference. . ." (p. 48).

Interference arising from a breakdown of inhibitory mechanisms appears to contribute to language comprehension problems (Hasher, Zacks, and May) by permitting the intrusion of irrelevant thoughts, personal preoccupations, and idiosyncratic associations. These irrelevant thoughts compete for processing resources, such as *working memory* capacity, and impair older adults' comprehension and recall. Hence, older adults' comprehension may be affected by distractions or intrusive thoughts. This hypothesis received support from a study by Kwong See and Ryan. Kwong See and Ryan examined individual differences in text processing attributable to working memory capacity, processing speed, and efficiency of inhibitory processes. Their analysis suggested that older adults' text processing difficulties can be attributed to slower processing and less efficient inhibition, rather than to working memory limitations.

Research by Connelly, Hasher, and Zacks compared passage reading times and answers to probe comprehension questions for young and older adults for texts that did or did not have distracting material interspersed amid target texts. The distractors, presented in a different type face, consisted of words or phrases conceptually related to the content of the target text and recurred over and over again throughout the target text. Connelly et al. reported that young adults not only read the texts containing the distracting material more rapidly than older adults but that they also showed greater comprehension of the target material. Connelly and colleagues' conclusion has been challenged by Dywan and Murphy who modified the procedure to include a surprise word recognition test for the interposed material. They found that the young adults had superior recognition memory for the distractor words, a result that is difficult to explain if the young adults are assumed to have been successful at inhibiting processing of the distractors. Burke also argues that research on the activation of word meanings and the detection of ambiguity provides "no support" for claims that "older adults are deficient in suppressing contextually irrelevant meaning or that they activate more irrelevant semantic information than young adults or that they retrieve more high frequency, dominant, or typical information than young adults" (p. P257).

Strategy differences may also underlie other age differences in language comprehension by affecting how readers process individual words. In general, young and older adults have been found to use similar reading strategies; however, age differences in reading strategies have been reported for specific aspects of syntactic and semantic processing. Stine found that young and older adults allocate reading time in similar ways to word-level and phrase-level processing. However, she also found young adults spent extra time reading words that occurred at sentence boundaries, minor clause boundaries, and major clause boundaries. While older adults also allocated extra time to major and minor clause boundaries, they did not spend extra time at sentence boundaries, suggesting older adults spend less time on sentence-level integration than young adults. Stine-Morrow, Loveless, and Soederberg (1996) let young and older adults read syntactically coherent text at their own pace. Both young and older adults who achieved good recall allocated extra reading time to syntactically complex sentences. However, some age differences were found with regard to other time allocation strategies used to achieve good recall. For young adults, good recall was related to the allocation of additional reading time to infrequent words and to new concepts first mentioned in the text. In contrast, for older adults, good recall was related to the allocation of additional reading time as they progressed serially through the text. These findings indicate that older adults use a different strategy than young adults to achieve good recall. Whereas young adults rely on recalling key words and concepts, older adults may rely on recalling a global text structure that is built up serially.

Despite working memory limitations, inhibitory deficits, and strategy differences, many older adults comprehend spoken and written language proficiently in everyday life. The age-related deficits observed in language comprehension studies may be offset by the ability to fill in missing elements of the discourse with meaningful reconstructions based on background knowledge and everyday reasoning abilities. Speakers and writers may also be able to minimize comprehension problems by using a special speech register, sometimes termed *elderspeak.* Elderspeak uses exaggerated pitch and intonation, simplified grammar, limited vocabulary, and slow rate of delivery. However, the use of elderspeak is controversial. On one hand, elder-speak may benefit older adults by reducing memory and processing demands. On the other hand, it may reinforce negative stereotypes of older adults and contribute to the social isolation and cognitive decline of older adults because it resembles "baby talk." Addressing older adults in "baby talk" by using short, simple sentences delivered very slowly and loudly with contrastive pitch seems to convey the impression to older adults that they are cognitively impaired and have communication problems (Kemper and Harden). Hence, practical techniques for modifying speech and writing targeted at older adults must reduce processing demands without triggering negative stereotypes.

C. Methodological Issues

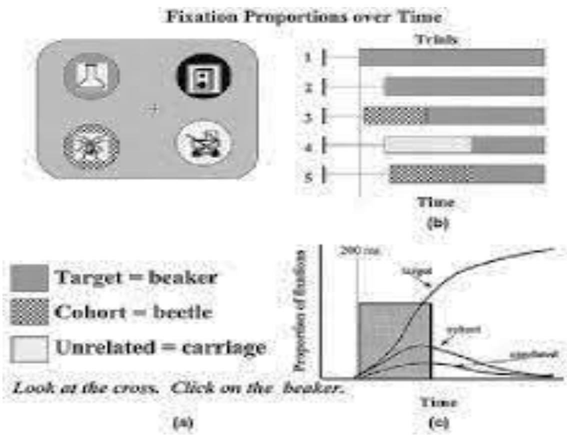
These early studies have raised numerous methodological questions, many of which were highlighted by the authors themselves. We now review what we see as the most important of these issues.

1. Data Analysis and Linking Assumptions

We will use Experiment 1 from Allopenna et al. (1998) to briefly describe how eyemovement data are analyzed. This experiment will also prove useful later for discussing some of the methodological concerns that arise in visual world studies in language comprehension. Allopenna et al. (1998) evaluated the time course of activation for lexical competitors that were cohorts, that is, they shared initial phonemes with the target word (e.g., beaker and beetle) or that rhymed with the target word (e.g., beaker and speaker). Participants were instructed to fixate a central cross and then followed a spoken instruction to move one of four objects displayed on a computer screen with the computer mouse (e.g., Look at the cross. Pick up the beaker. Now put it above the square).

2. Data analysis

A schematic of a sample display of pictures is presented in Figure 1 (Panel A). The pictures include the target (the beaker), a cohort (the beetle), a rhyme (speaker), and an unrelated picture (the carriage). The particular pictures displayed are used to exemplify types of conditions and are not repeated across trials. For current purposes,we restrict our attention to the target, cohort, and unrelated pictures. Panel B shows five hypothetical trials. The 0 ms point indicates the onset of the spoken word beaker. The dotted line begins at about 200 ms–the earliest point where we would expect to see signal-driven fixations, give the 150–200 ms required to program and launch a saccade (Matin, Shao, & Boff, 1993).



On the first trial, the hypothetical participant initiated a fixation to the target about 200 ms after the onset of the word, and continued to fixate on it (typically until the hand brings the mouse onto the target). On the second trial, the fixation to the target begins a bit later. On the third trial, the first fixation is to the cohort, followed by a fixation to the target. On the fourth trial, the first fixation is to the unrelated picture. The fifth trial shows another trial where the initial fixation is to the cohort. Panel C illustrates the proportion of fixations over time for the target, cohort, and unrelated pictures, averaged across trials and participants. These fixation proportions are obtained by determining the proportion of looks to the alternative pictures at a given time slice and they show how the pattern of fixations change as the utterance unfolds. The fixations do not sum to 1.0 as the word is initially unfolding because participants are often still looking at the fixation cross.

Although proportion of fixation curves might seem to imply that eye movements provide a continuous measure it is more accurate to say that eye movements can provide an approximation to a continuous measure. The assumption linking fixations to continuous word recognition processes is that as the instruction unfolds the probability that the listener's attention will shift to a potential referent of a referring expression increases with the activation (evidence for) of its lexical representation, with a saccadic eye movement typically following a shift in visual attention to the region in space where attention has moved. Because saccades are rapid, low-cost, low-threshold responses, a small proportion of saccades will be generated by even small increases in activation, with the likelihood of a saccade increasing as activation increases. Thus, while each saccade is a discrete event, the probabilistic nature of saccades ensures that with sufficient numbers of observations, the results will begin to approximate a continuous measure (see Spivey, Grosjean, & Knoblich, 2005; Magnuson, 2005).

A window of interest is often defined, as illustrated by the rectangle in Panel C. For example, one might want to focus on the fixations to the target and cohort in the region from 200 ms after the onset of the spoken word to the point in the speech stream where disambiguating phonetic information first arrives. The proportion of fixations to pictures or objects and the time spent fixating on the alternative pictures (essentially the area under the curve, which is a simple transformation of proportion of fixations) can then be analyzed. Because each fixation is likely to be 150–250 ms, the proportion of fixations in different time windows is not independent. One way of increasing the independence is to restrict the analysis to the proportion of new saccades generated to pictures within a region of interest. In the future, it will be important for psycholinguists to explore more sophisticated statistical methods for dealing with the temporal dependencies associated with how the linguistic input at time t effects location of fixations at subsequent temporal intervals.

Figure 2 (Panel A) shows the data from the Allopenna et al. (1998) experiment. The figure plots the proportion of fixations to the target, cohort, rhyme and unrelated picture. Until 200 ms, nearly all of the fixations are on the fixation cross. These fixations are not shown. The first fixations to pictures begin at about 200 ms after the onset of the target word. These fixations are equally distributed between the target and the cohort. These fixations are remarkably time-locked to the utterance: input-driven fixations occurring. 200–250 ms after the onset of the word are most likely programmed in response to information from the first 50 to 75 ms of the speech signal. At about 400 ms after the onset of the spoken word, the proportion of fixations to the target began to diverge from the proportion of fixations to the cohort. Subsequent research has established that cohorts and targets diverge 200 ms after the first phonetic input that provides probabilistic evidence favoring the target, including coarticulatory information in vowels (Dahan, Magnuson, Tanenhaus, & Hogan, 2001b, Dahan & Tanenhaus, 2004).

Shortly after fixations to the target and cohort begin to rise, fixations to rhymes begin to increase relative to the proportion of fixations to the unrelated picture. This result supports continuous mapping models, such as TRACE (McClelland & Elman, 1986), which predict competition from similar words that mismatch at onset (e.g., rhymes), but is inconsistent with the cohort model of spoken word recognition and its descendents (e.g., Marslen-Wilson, 1987, 1990, 1993), which assume that any featural mismatch at the onset of a word is sufficient to strongly inhibit a lexical candidate.

3. Formalizing a linking hypothesis

The assumption providing the link between word recognition and eye movements is that the activation of the name of a picture determines the probability that a subject will shift attention to that picture and thus make a saccadic eye movement to fixate it. Allopenna et al. formalized this linking hypothesis by converting activations generated by a TRACE simulation into response strength, following the procedures outlined in Luce (1959). The Luce choice rule is then used to convert the response strengths into response probabilities.

The Luce choice rule assumes that each response is equally probable when there is no information. Thus when the initial instruction is “look at the cross” or “look at picture X,” the response probabilities are scaled to be proportional to the amount of activation at each time step. Thus the predicted fixation probability is determined both by the amount of evidence for an alternative and the amount of evidence for that alternative compared to the other possible alternatives. Finally, a 200 ms delay is introduced because programming an eye movement takes 200 ms (Matin et al., 1993). In experiments without explicit instructions to fixate on a particular picture, initial fixations are randomly distributed among the pictures. Under these conditions, the simple form of the choice rule can be used (see Dahan et al., 2001a, 2001b). Note that the Allopenna et al. formalization is only an approximation to what would be a more accurate formalization of the linking hypothesis which would predict the probability that a saccade would be generated at a particular point in time, contingent upon (a) the location of the previous fixation (and perhaps the several preceding fixations; (b) time from the onset of the last fixation and (c) the current goal state of the listener's task– which can be ignored in a simple "click" task like the Allopenna et al. paradigm.

When the linking hypothesis is applied to TRACE simulations of activations for the stimuli used by Allopenna et al., it generates the predicted fixations over time shown in Figure 2 (Panel B). The predictions for the target, the cohort competitor, and a rhyme competitor closely match the behavioral data.

4. Action-contingent analyses

One useful feature of combining eye movements with an action is that the behavioral responses reveal the participant's interpretation. This allows for interpretation-contingent analyses in which fixations are analyzed separately for trials on which participants choose a particular interpretation. Two recent applications, illustrate how interpretationcontingent analyses can be used to distinguish among competing hypotheses.

McMurray et al. (2002) used a variation on the Allopenna et al. task to investigate the hypothesis that lexical processing is sensitive to small-within category differences in Voice-Onset Time (VOT). The stimuli were synthesized minimal pairs that differed only in voicing, such as bomb/palm and peach/beach. VOT varied in 5 ms step sizes from 0 to 40 ms. McMurray et al. found gradient increases in looks to the cross-category competitor as the VOT moved closer to the category boundary. While these results are consistent with the hypothesis that lexical processing is sensitive to within category variation, the results could also be accounted for without abandoning the traditional assumption that within-category variation is quickly discarded by making the following plausible assumption that there is noise in the system. For example, assume a category boundary of 18 ms. For trials with a VOT of 20 ms, given some noise, perhaps 20% of the stimuli might be perceived as having a VOT of 18 ms. With a VOT of 25 ms, the percentage might drop to

12%, compared to 8% for trials with a VOT of 30 ms and 4% for a VOT of 35 ms, etc. Thus, the proportion of looks to the cross-category competitor might increase as VOT approaches the category boundary because the data will include more trials where the target word was misheard as the cross-category competitor and not because the underlying system responds in a gradient manner.

McMurray et al. were able to rule out this alternative explanation by filtering any trials where the participant clicked on the cross-category picture. For example, if the VOT was 25 ms, and the participant clicked on the picture of the bomb, rather than the palm, then the eye-movement data from that trial would be excluded from the analyses. McMurray et al. found that looks to the cross-category competitor increased as VOT approached the category boundary, even when all “incorrect” responses were excluded from the analyses, thus providing strong evidence that the system is indeed gradient.

A second illustration comes from recent studies by Runner and his colleagues (e.g., Runner, Sussman, & Tanenhaus, 2003, in press) investigating the interpretation of reflexives and pronouns in so-called picture noun phrases with possessors, e.g., Harry admired Ken’s picture of him/himself. Participants were seated in front of a display containing three male dolls, Ken, Joe, and Harry, each with distinct facial features. Digitized pictures of the doll’s faces were mounted in a column on a board directly above each individual doll. The participant was told that each doll “owned” the set of pictures directly above him; that is,the three pictures in the column above Joe were Joe’s pictures, the pictures in the column above Ken were Ken’s pictures, etc.

Binding theory predicts that the reflexive, himself, will be interpreted as referring to Ken’s picture of Ken in instructions such as Pick up Harry. Now have Harry touch Ken’s picture of himself. Runner et al. found that looks to both the binding-appropriate and inappropriate referents began to increase compared to an unrelated picture in the same row, beginning about 200 ms after the onset of the reflexive. This result suggests that both binding-appropriate and inappropriate referents are initially considered as potential referents for a reflexive. However, participant’s choices showed frequent violations of classic binding for reflexives: on 20% of trials with reflexives, participants had Harry touch Ken’s picture of Harry. Thus, one might argue that the early looks to binding-inappropriate referents came from just those trials on which the participant arrived at the “incorrect” interpretation. Runner et al. were able to rule out this interpretation by analyzing just those trials where the participant made the binding-appropriate response, finding that there was still an increase in looks to the inappropriate referent compared to controls.

5. Task Variables

As the eye-movement literature on spoken language comprehension has developed, researchers have begun to vary the sorts of tasks given to their participants. The effects of these variations is important to evaluate and track from experiment to experiment since as discussed in the opening of this chapter, eye movement patterns are heavily task and goaldependent (i.e., we shift our attention to task-relevant regions of the world). It would be a mistake for instance, to assume that the “task” involved in the studies discussed in this chapter can be monolithically described as “spoken language comprehension” or worse still “use of language.”Very similar issues of task variation arise in reading eye-movement studies; eye-movement patterns over identical sequences of text will differ substantially depending on whether readers are skimming, understanding, memorizing, or proofing. Much greater opportunity for task variability appears to be possible in visual world studies because of the wide range of ways that participants can be asked to interact with the world. However, it is precisely this variability that provides experimenters with the leverage to make the visual world paradigm useful for such a wide range of questions.

One important task dimension is whether or not the linguistic stimuli used in the study involve instructions to act on the world. This variable is likely to be crucial because eye fixation plays an important role in visually guided reaching (see Hayhoe & Ballard,2005). At one extreme, imperative sentences are commonly used, such that participants are required to manipulate the objects (e.g.,Pick up the ball. Put it inside the cup.) At the other extreme, participants listen to declarative sentences, while looking at visually co-present referents. Here, the reference is intended to be non-deictic. (The boy picked up the ball. Then he put it inside the cup.)

Action-based studies offer several advantages in that participants are required in a highly natural way to remain engaged with their referent world; planning to execute a response requires calculating the spatial location of referents and presumably increases the time-locked nature of the relationship between linguistic interpretation and eye fixation. One clear limitation of the action-based paradigm however is that the linguistic stimuli must be embedded in instructions, which can limit the experimenter's degrees of freedom. The non-action-based listening procedure places far fewer constraints on both the experimenter and the participant. Decoupling fixations from action planning may also increase the proportion of anticipatory eye movements, which are extremely useful for inferring expectations generated by the listener.

Indeed, many of the most important applications of non-action-based listening have explored and documented referential expectations, starting with research initiated by Altmann and colleagues who showed that listeners can anticipate upcoming reference based on the semantic requirements of verbs and/or whole predicates (e.g., Altmann & Kamide, 1999; Kamide et al., 2003). Studies building upon this on this work include Boland (2005), who compared verb-based expectations for adjuncts and arguments, and Knoeferle and Crocker (in press) who studied the effects of visually based information on expectation about thematic role assignment.

We should note that this non-action paradigm is sometimes referred to as "passive" listening, and some investigators (e.g., Boland, 2005) have proposed that differences between fixations in action and passive listening tasks might be used to separate fixations that are controlled by language from those that are controlled by action. We are skeptical for several reasons. First, it is becoming increasingly clear that perception and action are inextricably intertwined in most perceptual domains, and we expect that this is also likely to be case for language. Second, interpreting sequences of fixations in the absence of an explicit task are likely to prove problematic for reasons eloquently articulated by Viviani (1990). We note however that many non-action task studies provide listeners with a welldefined task, typically so as to increase engagement with the scene and decrease the variability. For instance, Kaiser and Trueswell (2004) and Arnold et al. (2000) asked listeners to judge whether the depicted image on a trial matched the spoken description/story.

More generally it is important to keep in mind the following considerations. First, all saccadic eye movements involve some attentional overhead (Kowler, 1995). Second, the concept of passive listening leaves the underlying goals of the listener up to the listener.

Thus, each listener may adopt different goals, or worse, all listeners might adopt a pragmatically appropriate goal that was unforeseen by the experimenter. In short, there is no such thing as a taskless task. We therefore consider the notion of passive listening as akin to the notion of the null context, which is problematic for reasons articulated by Crain and Steedman (1985) and Altmann and Steedman (1988). Third, and perhaps most importantly, the difference between action-based (or perhaps more appropriately manipulation-based) and non-action-based variants of the visual world paradigm is really a subset of a more general question about the goal structures that control the moment-bymoment attentional state of the participants. In tasks with complex goal structures, e.g., a task-oriented dialog,multiple layers of goals will contribute to fixations,some of which may be are tied to expectations about upcoming linguistic input, some to the current subgoal, and some to higher-level planning.

Few studies to date have directly compared the action and non-action-based versions of the paradigm with the same materials (but cf., Sussman, 2006). However, to a first approximation, it appears that when anticipatory eye movements are excluded, the timing of fixations to potential referents may be slightly delayed in listening tasks compared to action-based tasks. The data from simple action-based tasks with imperatives (tasks where participants follow a sequence of instructions) is also somewhat cleaner than the data from non-actions-based tasks with declaratives, most likely because a higher proportion of the fixations are likely to be task-relevant.

6. Comparing Visual World and Eye-Movement Reading Studies

Many of the issues that have been investigated for decades using eye movements in reading, in particular issues in lexical processing and sentence processing are now being investigated using eye movements with spoken language. Although, some aspects of these processes will differ in reading and spoken language because of intrinsic differences between the two modalities, psycholinguists investigating issues such as syntactic ambiguity resolution and reference resolution using eye movements in reading and eye movements in spoken language believe they are testing theoretical claims about these processes that transcend the modality of the input. Thus, the psycholinguistic community will increasingly be faced with questions about how to integrate results from visual world studies with results from studies of eye movements in reading and sometimes how to reconcile conflicting results.

7. Processing load versus representational measures

In comparing reading studies to visual world studies it is useful to make a distinction between behavioral measures of language processing that measure processing difficulty and measures that probe representations. The distinction is more of a heuristic than a categorical distinction because many response measures combine aspects of both. Processing load measures assess transient changes in process complexity, and then use these changes to make inferences about the underlying processes and representations. Representational measures examine when during processing a particular type of epresentations emerges and then use that information to draw inferences about the underlying processes and representations. Neither class of measure nor its accompanying experimental logic is intrinsically preferable to the other; the nature of the question under investigation determines which type of response measure is more appropriate.

The majority of studies that use eye movements to examine reading make use of eye movements as a processing load measure. The primary dependent measure is fixation duration. The linking hypothesis between fixation duration and underlying processes is that reading times increase when processing becomes more difficult. In contrast, the majority of visual world studies use eye movements as a representational measure. The primary dependent measure is when and where people fixate as the utterance unfolds. We can illustrate these differences by comparing reading studies of lexical and syntactic ambiguity resolution with visual world studies that address the same issues.

8. Lexical ambiguity

In a well-known series of studies, Rayner and colleagues (e.g., Duffy, Morris, & Rayner,1988) have examined whether multiple senses of homographs,such as bank,ball, and port are accessed during reading, and if so, what are the effects of prior context and the frequency with each sense is used. Processing difficulty compared to an appropriate control is used to infer how ambiguous words are accessed and processed. For 'balanced' homographs with two more or less equally frequent senses, fixation duration is longer compared to frequency-matched controls–resulting in the inference that the multiple senses are competing with one another. This ambiguity "penalty"is reduced or eliminated for biased homographs when a 'dominant'sense is far more frequent than a 'subordinate' sense and when the context strongly favors either one of two equally frequent senses or the more frequent sense. Note that while these results do not provide clear evidence about time course per se, the overall data pattern allows one to infer that multiple senses are accessed, with the dominant sense accessed more rapidly. One can get crude time-course information by separately analyzing the duration of the initial fixation and using that as a measure of relatively early processes. More detailed information about time course can be obtained by using fixation duration as a measure, but using variations on the fast priming methods, introduced by Sereno and Rayner (1992).

A study using the visual world paradigm would adopt a similar approach to that used by Allopenna et al. Potential referents associated with the alternative senses would be displayed and the time course of looks to these referents would be used to infer degree of activation and how it changes over time. For balanced homophones, one would predict looks to the referents of both senses. For biased homophones, looks to the more frequent would begin earlier than looks to the less frequent sense. This pattern would be similar to those obtained in classic studies using cross-modal priming from the 1970s and early 1980s (Swinney, 1979; Tanenhaus, Leiman, & Seidenberg, 1979; for review see Simpson, 1984; Lucas, 1999). Note that these results would not provide direct information about processing difficulty, though one might infer from them that competing senses would result in an increase in complexity. Thus, while the eye-movement reading studies do not provide direct information about time course and visual world studies do not provide direct information about processing difficulty, the results from reading studies that use a processing load strategy and visual world studies that probe emerging representations could converge on the same conclusions.

9. Syntactic ambiguity

Beginning with the classic article by Frazier and Rayner (1982), eye tracking in reading has been the response measure of choice for psycholinguists interested in syntactic processing. Frazier and Rayner's approach was to examine the processing of temporarily ambiguous sentences,using reading times within pre-defined regions to infer if and when the reader had initially pursued the incorrect interpretation. For a range of syntactic ambiguities, most of which involved disambiguating the phrase that could be "attached" to a verb phrase, thereby introducing an argument, in favor of a noun phrase attachment that modified the head noun, Frazier and Rayner found an increase in fixation duration and an increase in regressive eye movements from the disambiguating region. For current purposes we will focus on fixation duration because it is most clearly a processing load measure.

The question of how to interpret regressions is more complex and beyond the scope of this chapter. The increase in fixation duration was interpreted as evidence that processing had been disrupted, thereby leading to the inference that readers had initially chosen the argument interpretation. Frazier and Rayner also introduced several different measures that divided fixations within a region in different ways. For example, 'first pass'reading times include all fixations beginning with the first fixation within a region until a fixation that leaves a region, and are often used as a measure of early processing.

Timing is less straightforward in eye-tracking reading when fixations are divided into multiple word regions. Most of the complexities in inferring time course in reading studies arise because the sequence of fixations need not correspond to the linear order of the words in the text. This is especially the case when one considers that arguments about timing often depend on defining regions of text and then partitioning fixations into categories in ways that separate the measure from when the input is first encountered.

Studies examining syntactic ambiguity resolution with the visual world paradigm use the timing of looks to potential referents to infer, if and, if so, when, a particular analysis is under consideration. That one can align fixations and the input is, of course, because the input unfolds sequentially. Note, however, that one cannot use fixations in a straightforward way to draw inferences about processing difficulty. Thus the visual world approach is unlikely to become a paradigm of choice for investigating issues about resource demands, including increasingly important questions about what factors contribute to the complexity of sentences (e.g., Grodner & Gibson, 2005; Hale, 2003; Lewis & Vasishth, 2005).

10. Effects of Display

The single factor that most complicates the interpretation of visual world studies of language processing is the need to use a display. First, the encoding of the display can introduce contingencies. For example, the timing of looks to a potential referent at point t could be affected by whether or not that referent has been fixated on during time t-x, either during preview or as the sentence unfolds. Thus the likelihood of a fixation may be contingent on both the input and the pattern of prior fixations. This, of course, has the potential to complicate inferences about time course, in much the same way that re-reading after a regression can complicate the interpretation of fixation duration data in eyemovement reading studies. Recent studies have begun to examine how having fixated a potential referent during preview affects the likelihood that it will be fixated when it is temporarily consistent with the input (Dahan, Tanenhaus, & Salverda, in press).

Second, use of a display with a small number of pictured referents or objects and a limited set of potential actions creates a more restricted environment than language processing in most natural contexts, while at the same time imposing more demands on the participant than most psycholinguistic tasks. In order to address these closed set issues, we will consider two cases: the first from spoken word recognition; the second from reference resolution.

11. Spoken word recognition

In the Allopenna et al. paradigm, the potential response set on each trial is limited to four pictured items. If participants adopted a task-specific verification strategy, such as implicitly naming the pictures, then the unfolding input might be evaluated against these activated names, effectively bypassing the usual activation process, and leading to distorted results. Even if participants do not adopt such a strategy, the visual world methodology might be limited if the effects of the response alternatives mask effects of non-displayed alternatives (e.g., neighborhood effects in the entire lexicon). This would restrict its usefulness for investigating many issues in spoken word recognition, in particular issues about the effects of lexical neighborhoods, i.e., the set of words in the lexicon that are similar to the target word. Here, an analogy might be helpful. Researchers often use lexical priming paradigms to probe for whether an exemplar of a particular class of lexical competitor is active, for example, cohorts or rhymes. However, these paradigms are not well suited for asking questions about the aggregate effects of the number and frequency of potential competitors. In order to investigate this class of question, researchers have found it more useful to measure response time to a target word, for example, auditory lexical decision, which more closely approximates a processing load measure.

12. Implicit naming

The issue of implicit naming has been addressed most directly by Dahan and Tanenhaus (2005) in a study that varied the amount of preview time, 300 or 1000 ms, for four-picture displays with minimal phonological overlap between the names of the distractors and the target one or both words. For example, the pair snake–rope was selected because the picture of a coiled rope shares some features with the visual representation most often associated with the concept of a snake. When selecting pictures, Dahn and Tanenhaus (2005) sought to minimize their visual similarity so that the objects could be easily differentiated. For example, we chose a snake in a non-coiled position. Thus, visual similarity was maximized between the prototypical visual representation of one of the concepts, the referent, and the picture associated with the other concept, the competitor, and minimized between the competitor picture and the picture of the referent concept.

Several aspects of the results provide strong evidence against implicit naming. Preview duration did not affect the magnitude of visual similarity effects (looks to visually similar competitors). Moreover, even in the 1000 ms condition, the magnitude of visual similarity effects was not affected by whether or not the competitor was fixated during preview; the naming hypothesis predicts that effects would be eliminated or weakened with preview because the encoded name of the picture would not match the unfolding target. Finally, similarity effects were larger when the target had a competitor that was chosen to share visual features of its prototype representation compared to when that competitor was the referent. Thus visual similarity effects were due to the fit between the picture and the conceptual representation of the picture,not simply surface visual confusability. This suggests that mapping of the word onto its referent picture is mediated by a visual/conceptual match between the activated lexical form of the target and the picture. This hypothesis is further supported by additional analyses of the effects of fixation to a competitor during preview on the likelihood that it will be re-fixated during the speech input and evidence that a spoken word triggers looks to potential referents when the participant is engaged in a visual search task to identify the location of a dot when it appears on a random location within a schematic scene (Salverda & Altmann, 2005).

13. Sensitivity to hidden competitors

Perhaps, the strongest test of the sensitivity of visual world studies comes from studies that look for effects of non-displayed or “hidden competitors.” For example, Magnuson, Dixon, Tanenhaus, and Aslin (in press) examined the temporal dynamics of neighborhood effects using two different metrics: neighborhood density, a frequency-weighted measure defined by the Neighborhood Activation Model (NAM),and a frequency-weighted measure of cohort density. The referent was displayed along with three semantically unrelated pictures, with names that had little phonological overlap with the referent (all names were monosyllabic). Crucially, none of the referent’s neighbors were either displayed or named throughout the course of the experiment. The results showed clear effects of both cohort and neighborhood density, with cohort density effects dominating early in the recognition process and neighborhood effects emerging relatively late.

These results demonstrate that the processing neighborhood for a word changes dynamically as the word unfolds. It also establishes the sensitivity of the paradigm to the entire lexicon. To a first approximation then, when competitors are displayed, the paradigm can be used to probe specific representations, however, the aggregate effects of competitors can be observed in the timing of fixations to the target referent.

Magnuson et al.'s results complement Dahan et al. (2001b) finding that misleading coarticulatory information delays recognition more when it renders the input temporarily consistent with a (non-displayed) word, compared to when it does not. In addition, simulations using the Allopenna et al. linking hypothesis successfully captured differences between the effects of misleading coarticulatory information with displayed and non-displayed competitors. Whether the non-displayed competitor logic can be extended to higher-level sentence processing remains to be seen.

14. Sentence processing

Much trickier issues about the effects of the display come into play in higher-level processing. For example, one could argue that in the Tanenhaus et al. (1995) study displaying an apple on a towel and an apple on a napkin increases the salience of a normally less accessible sense compared to circumstances where the alternative referents are introduced linguistically. One could make a similar argument about the effects of action on the rapidity with which object-based affordances influence ambiguity resolution in studies by Chambers and colleagues (Chambers, Tanenhaus, Eberhard, Filip, & Carlson, 2002; Chambers et al. 2004). In these studies, the issue of implicit naming seems prima facie to be less plausible. However, one might be concerned about task-specific strategies. For example, in Chambers et al. (2002), participants were confused, as indexed by fixations when they were told to, Pick up the cube. Now put the cube in the can, and there were two cans. The confusion was reduced or eliminated, however, when the cube would only fit in one of the cans. Because only one action was possible, one might attribute this to problem solving, and not as Chambers et al. argued to the effects of action and affordance on referential domains. However, the manipulation had opposite effects for instructions that used an indefinite article, e.g., Pick up the cube. Now put it in a can. Here participants were confused when the cube would only fit in one of the cans. This strategy of pitting linguistic effects against potential problem-solving strategies is crucial for evaluating the impact of strategies due to the display and the task.

Perhaps, the most general caution for researchers using the visual world paradigm in both production and comprehension is to be aware that while the visual world displays entities that can be used to infer the representations that the listener is developing, it also serves as a context for the utterance itself. Note that the fact that information in a display affects processing is not itself any more problematic than the observation that reference resolution, for example, is affected by whether or not potential referents are introduced linguistically in a prior discourse. One sometimes encounters the argument that the visual world paradigm can be informative about language processing only if gaze patterns to a potential referent in a display are not affected by the other characteristics of the display. This argument is no more or less valid than the comparable argument that fixations in reading can only inform us about word recognition or reference resolution if fixations to a word are unaffected by the context in which the fixated word occurs. What is crucial, however, is whether the nature of the interactions with the display shed light on linguistic processing or whether they introduce strategies that mislead or obscure the underlying processes. Thus, far investigations of potential problems has been encouraging for the approach. However, it will be crucial in further work to explore the nature of the interactions between the display and linguistic processing in much greater detail.

CONCLUSION

This chapter has provided an overview to the rapidly growing literature on eye movements and spoken language processing, focusing on applications to spoken language comprehension. We have reviewed some of the foundational studies, discussed issues of data analysis and interpretation, and discussed issues that arise in comparing eye-movement reading studies to visual world studies. We have also reviewed some of the major lines of research that are utilizing this method, focusing on topics in language comprehension, including spoken word recognition, use of referential constraints in parsing, interactive conversation, and the development of language processing abilities in children.

It should be clear from this review that the visual world paradigm is being employed in most traditional areas of inquiry within psycholinguistics. And in each of these areas, the visual world approach is encouraging psycholinguists to investigate uncharted theoretical and empirical issues. Within the study of spoken sentence comprehension, issues about reference have taken center stage, in part because the visual world methodology makes it possible to connect research on real-time reference resolution with social and cognitive research on pragmatics and conversation. Within the study of spoken word recognition, the time-locked nature of this measure has allowed researchers to explore phonemic and sub-phonemic and prosodic contributions of word recognition in utterances at a level of detail previously not possible with traditional methods. It is for these reasons and other reasons we are quite optimistic that eye-movement measures will continue rise in interest and use within the psycholinguistics community.

We close by noting that eye-movement measures are likely to be most powerful when combined with other measures. We have seen how combining eye movements with action and structure tasks can shed new light on real-time language processes. We expect that other measures will emerge that provide additional advantages. For instance, other body movements pertaining to gestures and actions are likely to be highly informative when connected to the timing of speech and eye gaze events. Most generally, we see the visual world approach as part of a larger movement toward connecting language and action in rich goal-directed tasks using increasingly rich and complex data arrays to understand the dynamics of comprehension and production in conversation. This approach is likely to have an increasingly important influence on theoretical development in natural language, just as it as it has begun to enrich theories in other areas of perception and cognition (Ballard, Hayhoe, Pook, & Rao, 1997; Barsalou, 1999; Hayhoe & Ballard, 2005; Land, 2004).

[This page is intentionally left Blank]

CHAPTER 14
PSYCHOLINGUISTICS AND NEUROLINGUISTIC PERSPECTIVES OF SIGN LANGUAGES

INTRODUCTION

By now most of us have had the opportunity to experience sign language, if only to see it occasionally in the corner of our TV screens. There we can see a person translating speech into sign for the benefit of deaf and severely hearing-impaired viewers. You may wonder whether those signs truly are part of a language or are just a collection of gestures that lack the sophistication of a language based on speech. How can we judge whether persons who use 'sign language' truly have language? First, a sign language is a true language because the language system allows a signer to comprehend and produce an indefinitely large number of grammatical sentences in signs. This feat can be accomplished with a limited number of signs (vocabulary) and a system (syntax and semantics). Secondly, a signing person has a true language if that person can communicate by sign whatever can be communicated by speech. This is reasonable because we can all agree that people who communicate in speech do have language.

There is, of course, a difference in the physical means of communication: signing involves light and speech involves sound. But a particular physical mode is not an essential aspect of language. Of course, language must depend on some physical mode for its use and learning but that mode need not be limited to sound. Thus, what we are really interested in here is, for example, whether an abstract sentence like 'If the weather had been fine, then Mary's uncle could have come and given her money' can be conveyed through signing. It can. Such a sentence is a good test, since it expresses a variety of complex semantic functions and relations and involves a number of events and situations, none of which, interestingly enough, had occurred. According to the sentence, the weather was not fine, Mary's uncle did not come, and the uncle did not give money to Mary. A person who could comprehend and produce communications such as this, through sign rather than speech, surely can be said to have language.

Research on sign languages seriously began for the most part in the 1960s when linguists and psycholinguists addressed themselves to this newly discovered area. The findings showed that signers of such sign languages as American Sign Language, French Sign Language, British Sign Language, and others, can indeed communicate in sign whatever is expressed in speech (Stokoe et al., 1965; Klima and Bellugi, 1979; and, more recently, Siple and Fischer, 1990). A sentence like that shown at the end of the previous section can be expressed through all of these languages. Other sign languages may be incomplete syntactically or limited in terms of vocabulary. Such incomplete sign languages are typically found in developing countries, although in even some developed nations, sign language may suffer from deficiencies. In Japan, for example, where the national government until recently prohibited the teaching and use of sign language in public schools, standardization and vocabulary are problems.

Not only can a fluent signer of a complete sign language such as American Sign Language (ASL) sign whatever a speaker can say, but the signer communicates at about the same speed as a speaker does. The speed at which signers produce sentences (more precisely the ideas that underlie sentences) in a signed conversation tends to be similar to that at which speakers produce sentences in a spoken conversation (Bellugi and Fischer, 1972). This occurs even though a signer, as does a speaker, has the ability to exceed this speed. There seems to be an optimum speed at which humans are comfortable in processing language information, whether that information be in the form of speech or sign.

There may even be strong dialectic differences within a language from region to region within a country. For example, signers from Paris have difficulty in understanding signers from Lyon, and vice versa. The reader may be surprised to know, too, that American Sign Language and British Sign Language (BSL) are not mutually intelligible. American Sign Language actually has more in common with French Sign Language than with British Sign Language because ASL was derived from French Sign Language early in the nineteenth century. British Sign Language had its beginnings before the seventeenth century, with the first deaf schools opening in Britain in the late eighteenth century.

This might be a good place to emphasize that, contrary to common belief; there is no universal sign language. There are some similarities among languages, but not many. There is a multitude of sign languages, complete and incomplete. If there were a campaign for an artificial sign language to be adopted by signers everywhere, it would run into the same opposition that Esperanto has met in its quest to serve as a universal speech-based language. Like speech-based languages, a sign language is part of a culture. It may be useful to learn a foreign (or second) language but to give up one's native language, be that speech or sign, is something that people are not eager to do.

A. Single Sign Recognition

As in spoken language psycholinguistics, much of the literature on sign language psycholinguistics has sought to uncover the structure of the lexicon, and the processes governing the access of lexical items stored therein. Most contemporary models of spoken and written word recognition conceptualize lexical access as a matching process between a perceptual signal that accrues over time, and potential lexical candidates stored in memory (e.g., Cohort model, Neighborhood Activation model and TRACE; for a review, see Jusczyk & Luce,2002). Competitive activation among lexical candidates ultimately yields a single word percept. Questions that are especially relevant to sign language psycholinguistics, therefore, include how the physical properties of the sign language signal affect the way in which the lexicon is initially accessed en route to word recognition, and what organizational principles dictate how signs are stored within the lexicon.

In spoken languages, word recognition is achieved through the mapping of the acoustic speech signal received by the ear onto word forms stored as mental representations in the listener's mind. The mapping process unfolds in time as the acoustic signal of the word accrues, and is thought to be mediated by psychological processes which, in part, serve to rectify variances in the signal due to articulatory factors (e.g., allophonic variation) and cross-speaker differences. The details of the initial phonetic mapping, the structure of the representations, and the structure of the mental lexicon have been elucidated by psycholinguistic research. Well-accepted findings include lexicality and word-frequency effects, context effects (including semantic- and form based priming), and neighborhood effects.

Lexically effects refer to the finding that words are recognized faster than non-words. This basic finding suggests that the presence of a mental representation (i.e., a known word) provides a target for a search though the lexicon, whereas the lack of a representation engenders an exhaustive and unfilled search and thus incurs a processing disadvantage. Word frequency effects suggest that words that are highly frequent in the language (e.g., BABY) are recognized faster than low-frequency words (e.g., DRAGON). Psycholinguists have proposed a threes holding account of these data, with highly frequent words having a lower resting threshold and thus require less information for recognition. Context effects may have a temporal component, Whereby a word that has been previously encountered may affect the subsequent processing of incoming words. For example, the recognition of the word CAT is faster if subjects previously heard the word PIG versus an unrelated word like PAIL. This robust effect is known as semantic priming. These data suggest that the lexicon is structured along a semantic dimension, such that semantic features may be shared between entries and thus may be co-activated, leading to speeded processing of related entries.

While only a few psycholinguistic studies of on-line processing in signed languages exist (see Emmorey, 2002 for a recent review of psycholinguistic studies of ASL), there is a growing body of literature confirming that factors affecting spoken word recognition also influence the recognition of signed language signs. For example, the importance of the lexical status of a sign form was demonstrated in a study in which participants were required to make a lexical decision about a target sign, after first being primed by a related sign or sign-form. Reaction times were significantly slower to formational possible but non-existing ASL signs than to real ASL signs (Corina & Emmorey, 1993). Similarly, Dye and Shih (in press) reported data from native users of British Sign Language which showed lexicality effects (with slower responses to non-signs) in that language.

B. From Signal to Representation

Although signs and spoken words are formational quite different, we have reason to believe that words in each modality unfold in time in a lawful fashion, and that accessing lexical representations may vary as a function of sub lexical properties. For example, in a gating task study by Emmorey and Corina (1990), signs were partitioned into 33 ms parts and presented to participants cumulatively. Participants identified the location of the sign first, followed quickly by the hand shape and finally the movement. Signs located in neutral space were recognized before those located upon the face, presumably because the target location of the sign was achieved earlier for neutral space signs. In addition, it was observed that signers could anticipate changes in signs that included a hand shape change prior to full articulation of the hand shape. This demonstrates that in sign recognition, as with spoken language, co-articulatory factors may assist word identification.

Finally, these studies confirmed that signs, though typically longer in duration than spoken words, were identified very quickly. In contrast to spoken English words in which approximately 83% of a word must be heard before identification occurs, sign identification occurred only after approximately 35% of the sign form had been seen. Two factors that may account for this finding include the relatively greater simultaneous packaging of phonological information within a sign and the fact that few signs share an initial phonological shape, leading to a reduced number of competing lexical items (i.e., reduced initial cohort size).

C. Form-Based Structure of the Lexicon

While the lexicon may be organized along a semantic dimension such that words sharing a meaning relation prime one another, there is also substantial support for structural properties of words influencing processing. For example, the processing of the word CAT can be affected by a previous encounter with a similar sounding word like MAT more than an unrelated word like LIP. This suggests a further dimension of lexical organization that honors a structural relationship between words (e.g., the phonological status of shared "rime").

In sign, frequency and semantic context speed recognition. In contrast, structural relations (e.g., phonological composition of words and lexical neighborhoods) have been reported to typically produce inhibitory effects. Lexical decision studies have sought to establish processing effects based upon shared formational similarity (i.e., signs sharing one or more parameters), but have produced discrepant findings. For example, Corina and Emmory (1993) reported inhibitory effects for targets sharing an articulatory location with primes, no effects for shared hand shapes and a facilitator effect for movement. More recently, Corina and Hildebrandt (2002) investigated movement and location priming at 500 and 100 ms inter stimulus interval (ISI) lags. They found no evidence of phonological priming for either Movement or Location at the 500 ISI lag, though inhibitory non-significant trends (Movement p 0.064, Location p 0.088) were observed both for location and movement at 100 ISI lag. In contrast, Dye and Shih (in press) found some evidence of facilitator phonological priming in native signers for signs that shared a common location and/or location and movement. Importantly, these effects were not observed when primes were non-signs, suggesting that the observed priming effect cannot be divorced from lexicality.

The inconsistent findings in structural priming studies of signed languages may reflect different sources of activation. The studies of Corina and Emmory (1993) and Dye and Shih (in press),which report facilitation, document unusually long reaction times for sign target detection (————————————1100 ms), while the study that reported inhibition has reaction times in the 800 ms range. There is a growing consensus in the psycholinguistic literature that these longer reaction times may reflect post-lexically mediated or controlled processing strategies, while inhibition represents the effects of automatic priming. That is, these inhibitory effects reflect lexical rather than post-lexical access.

A related and well-known factor influencing lexical access across both written and spoken language modalities is the composition of the candidate set of lexical entries from which a single target form must emerge. In spoken languages these effects are conceptualized as owing to competition among formational similar lexical entries—so-called "neighborhood effects." The metric of lexical neighborhood similarity has been traditionally defined in terms of phonological properties in studies of spoken words, and orthographic properties in studies of written words. For example, Luce and Pisoni (1998) derive neighborhood similarity by considering the number of words that could be obtained by a single phoneme substitution, addition, or deletion. Numerous studies have demonstrated that these similarity properties influence word recognition, as shown in the perceptual identification accuracy, naming and lexical decision latencies, and priming effects, both in spoken and visual word recognition (for discussion, see Jusczyk & Luce, 2002). The frequency of individual lexical items may also interact with properties of similarity. Importantly, the effects of similarity and word frequency have been observed across many languages. A general characteristic of these effects is that lexical access tends to be inhibited as similarity increases, especially when lexical frequency is held constant.

A series of studies conducted with Lengua Signos Español (LSE) have examined the interaction between sign similarity and frequency (Carreiras, Gutiérrez-Sigut, Baquero, & Corina, submitted). As sign frequency counts are unavailable for most signed languages, this measure was formalized as a familiar rating based upon how often a particular sign was used. Lexical similarity was evaluated using the Hamburg Sign Language Notation System (Prillwitz, Leven, Zienert, Hanke, & Henning, 2004), whereby the degree of similarity for the parameters of location and hand shape were calculated by counting how many signs were produced in a particular location or articulated with a specific hand shape. These studies reveal that for neighborhoods defined by location, low-frequency signs are recognized slower in high density neighborhoods compared to sparse neighborhoods. A different pattern was observed for neighborhoods defined by hand shape. Here, low-frequency signs were recognized faster in dense neighborhoods compared to sparse neighborhoods.

Several explanations for the processing differences for location and hand shape are offered. In many theoretical models of sign structure, hand shapes are represented as multi-featured, compositional and hierarchically ordered entities that serve as auto segments within skeleton-based phonological treatments (Corina & Sandler, 1993). Feature representations of location tend to be far less compositional, often adopting a unitary representation such as [shoulder] for a sign articulated on or near the shoulder (Brentari, 1998). These variations in richness of representation may differentially impact lexical activation.

It is interesting to note that in spoken word recognition there is some evidence for word onset information producing inhibitory effects (e.g., Luce & Pisoni, 1998; Vitevitch & Luce, 1998). As noted above in the gating studies of sign recognition, location was the first parameter to be accurately and consistently identified during the unfolding of a sign, followed by handshape. This is not to suggest that location serves as a syllabic onset in sign languages, but rather that there may be a special status of first recognized word/sign elements across languages.

D. Effects of Language Experience

A different approach to the study of the underlying structural properties of signs and the sign lexicon is through the study of phonological awareness in deaf signers. Phonological awareness reflects an understanding that words are not segmented wholes, but are made up of component parts. Some structural groupings in spoken languages have a special status that can often be observed in stylized language usage such as poetry, language games, and song. For example, words that rhyme (e.g., juice-moose) is generally appreciated as more similar-sounding than are words that share only onsets (e.g., juice-June).

Signed languages exhibit similar privileged groupings. Controlled studies reveal that similarity judgments of signs sharing two sign parameters (i.e., hand shape and movement) produced much more consistent judgments than signs sharing only one parameter (Hildebrandt & Corina, 2002). Both deaf and sign-naive subjects judge signs that share movement and location as the most similar, indicating that this combination of parameters enables a robust perceptual grouping. This fact accords well with the observation that languages commonly capitalize on perceptual distinctions as a basis for linguistic distinctions. For example, theories of the syllable structure of ASL have proposed that the combination of movement and location properties serves as the skeletal structure from which syllables are built, and that movement is the most sonorous element of the sign syllable (see, for example, Sandler, 1989). Interestingly, however, response patterns of the deaf native signers indicate that under some circumstances, tacit knowledge of ASL syllable structure may override purely perceptual factors.

The effects of language experience on the perception of signing are further illustrated in the studies investigating the phenomenon of categorical perception. In the spoken language domain, this phenomenon refers to the finding that speech stimuli are perceived categorically rather than continuously despite a continuous variation in form (Liberman, Cooper, Shankwelier, & Studdard-Kennedy, 1967). The principle of categorical perception helps explain how listeners may resolve the many-to-one mapping between continuous acoustic patterns and discrete phonological categories. Initially, the phenomenon of categorical perception was taken as evidence for a hard-wired, language-specific mechanism, but more recent work has placed the phenomenon within the framework of a general psychophysical mechanism that may be observed over a variety of non-language domains.

Studies of categorical perception for phonological parameters in ASL have been conducted for place of articulation, hand shape, and facial expressions in signers and non-signers. Recently, categorical perception was found for the parameter of hand shape, but not place of articulation (Emmorey, McCullough, & Brentari, 2003), an effect that was limited to deaf signers (see also Baker, Isardi, Golinkoff, & Petitto, 2005). This indicates that linguistic knowledge can influence perceptual judgments.

While it has been demonstrated that hearing non-signers exhibit categorical perception for emotional facial expressions (Calder,Young, Perret, Etcoff, & Rowland, 1996), the study of categorical perception for facial information is interesting in signed languages, as facial information in ASL not only conveys emotional information about the state of the speaker, but also serves a linguistic function. Several well-defined facial configurations serve to signal adverbial modulations and specify syntactic forms. Recent studies have found categorical perception of both emotional and linguistic facial expression (McCullough & Emmorey, 1999; Campbell,Woll, Benson, & Wallace, 1999). However, these effects were observed in both deaf signers and hearing non-signers, suggesting that categorical perception for facial expression is not mediated by linguistic knowledge per se.

E. Sign Language Production

As with the study of spoken languages, psychological processes underlying the production of sign language have been underrepresented in the literature. However, examples of “slips of the hand ”akin to “slips of the tongue, ”have been reported in signed languages. In these data, semantic- and form-based errors are attested, but dual (semantic

---

phonological) errors are rare. For example, in a report of production errors in German Sign Language (Deutsche Gebaerdensprache [DGS]), 38/40 sign-substitution errors are semantically based; only one is semantically

---

form-based (Hohenberger, Happ, & Leuninger, 2002). Sign language form-based errors manifest primarily as anticipations and perseverations. Interestingly, there is a marked disparity in the frequency with which the individual sub-lexical form parameters are subject to error, with hand configuration being far more susceptible to error than place of articulation and movement (Hohenberger et al., 2002). These data are consistent with previous slip literature (Klima & Bellugi, 1979; Newkirk, Klima, Pedersen, & Bellugi, 1980) in which the majority of form-based errors are in hand configuration.

Production errors akin to the “tip of the tongue” phenomenon have been observed sign languages. A recent report of the so-called “Tip of the Finger” (TOF) experiences in deaf signers indicate that, as is observed in studies of spoken language, proper names tend to invoke a large percentage of TOF states (Thompson, Emmorey, & Gollan, 2005). However, as proper names in ASL are nearly always finger spelled (rather than signed using a lexical sign), it cannot be assessed whether the contributions of TOF state is based upon a failure to retrieve of the English spelling of a proper name or its corresponding finger spelled form. TOF states do occur in lexical signs (though less frequently),and these forms provide an confounded means by which to assess the properties partial retrieval of lexical forms in signing. These studies indicate that a deaf subject in a TOF state has some access to the phonological composition of an otherwise inaccessible sign. Specifically, it appears that some information regarding the initial configuration of the hand shape, location, and orientation is more (and equivalently) reportable, relative to details of the sign movement. Thus similar to spoken language, word onsets appear to have a privileged status on word retrieval in signed languages. The apparent simultaneous accessibility of three of the four major phonological parameters of sign suggests that lexical retrieval may not be guided by a single phonological parameter.

Further evidence of this comes from a sign language production experiment (Corina & Knapp, in press) in which the time-course of semantic and phonological form retrieval in sign was assessed. In this paradigm, signers were required to articulate sign names of an object while observing an overlay of a distractor sign. This Stroop-like task found that native signers of ASL exhibit longer naming latencies in the presence of semantically related sign distractors than in the presence of unrelated distractors at early, but not late, stimulus onset asynchronies (SOAs). In contrast, phonologically related stimuli produce naming facilitation both early and late in the naming process, but effects vary by degree and type of phonological relatedness. Interestingly, we observed different amounts of interference between distractors that shared one, two, and three parameters with the target, with greatest effects observed when the sign target to be articulated and the interfering stimulus shared both movement and location parameters. Note that the combination of location and movement was an important factor in the lexical decision experiment of Dye and Shih (in press). As discussed previously, location and movement components of signs are linguistic categories that may comprise the skeletal structure of a sign.

F. Morphology

Morphological theories are formal statements about how word forms are built. Traditional approaches provide accounts of how new words are adopted into the existing lexicon (i.e., coinage), the relationships between word forms that change grammatical category and thus extend meaning usage (i.e., derivational morphology), and how word forms adapt in the face of syntactic phenomena (inflectional morphology).

Psycholinguistic studies of sign language morphology have been largely driven by the fact that morphological processes do not easily fit into a traditional segment-based approach. In contrast to spoken languages in which morphological forms are commonly created by prefixing or suffixing onto a stem, many of the morphological form changes observed in signed languages are non-concatenative. Under morphological inflection, the entire base form undergoes complex movement modulations.

Consider, for example, the morphological changes that affect a verb like TELL. In the citation form of this sign, the extended index finger of the otherwise closed hand shape touches the chin with the palm facing the signer. It moves forward in a straight path via the extension of the elbow to approximately a 120 degree angle. In one aspectual inflection that signals intensity or forcefulness, we find a reduction in elbow extension and an increase in movement velocity. A temporal inflection, giving rise to a meaning that this action occurred again and again over time, adopts an elliptical path movement that passes by the chin in repeated cycles. Inflection for grammatical number agreement may take on a sequence of repeated straight movements directed towards two or three laterally and evenly displaced points in space. Finally, grammatical verb agreement inflections modulate the beginning and end of the path movements which in turn signal the grammatical subject and object of the sentences, respectively.

Many psycholinguistic studies of spoken language morphology have examined whether morphologically complex forms are deconstructed into constituent parts or treated in a more holistic fashion. Current indications are that the degree of decomposionality may vary with the formal devices used to modulate word meaning in a specific language. The question of how complex sign forms are parsed has been dealt with in only a handful of studies. Emmorey (1991) investigated the organization of the lexicon for morphologically complex forms using a repetition priming experiment. Here, a morphologically inflected sign was used as a prime, and a lexical decision was made about its uninflected form. These studies found that whereas sign primes inflected for aspect (e.g., habitual, continual) did produce facilitation in the recognition of the associated uninflected stem, signs inflected for grammatical agreement (e.g., dual, multiple, reciprocal) did not prime their citation forms. Emmorey (2002) has suggested that differing degrees of productivity may affect the lexical representation of morphologically complex forms in ASL. The interplay between the numbers of verb types that participate in a particular inflectional process may influence the association between related forms in the lexicon. For example, more verb forms can be marked with the habitual inflection than the reciprocal inflection.

G. Syntax

Syntax refers to a level of language form that specifies the interrelationship between words in a sentence. Formal studies of syntax reveal universal and language-specific patterns and constraints that reliably contribute to interpretation of the sentence meaning, especially those specifying thematic roles. Psycholinguistic studies of syntactic processing strive to understand how these structured patterns are exploited during on-line comprehension of sentences, and how this information is combined with word meaning to give rise to a conceptualization of the intended message.

Early spoken language studies examined the psychological reality of clause boundaries and syntactic complexity (Garrett, Bever, & Fodor, 1966). More recent studies have attempted to document the precedence of sentence interpretation as it unfolds in time, with some suggesting a leading role of syntactic over semantic properties. Other studies have examined how structural dependencies in a sentence are instantiated – for example, the integrative processes by which a noun and its associated pronoun are coupled in sentences like “John went to the store and he bought some bread.” The few studies of syntactic processing in signed languages have focused on the phenomenon of co-reference, exploiting the unusual means by which pronouns and their noun antecedents are associated. Specifically in signed languages like ASL, a signer may associate a noun with a specific location in articulatory space (typically on a horizontal plane in front of the signer) with a later indexes point to this space, signaling a co-referential relationship. Thus in the sentence above, the proper noun John may be articulated in a right-sided location in neural space, a subsequent point to that same location would signal the antecedent “John.” In many theories of sign language grammar this pointing sign is considered a pronoun.

Spoken language studies have found that at the time one encounters a pronoun in a sentence one can find evidence that its noun antecedent has been “reactivated.” Using a probe recognition technique, Emmorey and colleagues found evidence for antecedent re-activation during sign comprehension that was similar to that observed for spoken languages. A semantically related probe sign that followed the pronoun was recognized faster than a semantically inconsistent sign. The authors also explored several sign language-specific aspects of co-reference. For example, the re-activation was noted even when the grammatical co-reference was signaled by verb agreement rather than by an overt pronoun, thus providing a psycholinguistic validation of the theoretically motivated analysis of null-pronoun phenomena in ASL (Emmorey & Lillo-Martin, 1995). In addition, Emmorey, Corina, and Bellugi (1995) examined whether the spatial location of the probe interacted with the entailed semantic relationships. They found that the consistency of the spatial location of the probe item did not influence response times. That is, while a semantically (and co-referentially) appropriate probe sign was recognized faster than a semantically inconsistent sign, these recognition times were not modulated by absolute spatial location of the probe (which could appear in the same location as the pronoun [and antecedent] or a different location). Interestingly, this lack of spatial effect was observed only in sentences in which the spatial locations signaled grammatical relations. In contrast, when the specialization of the nouns and pronouns made reference to actual topographic relationships (i.e., real-world space, such as “chair located to the left of a table”) the consistency between the probes and the spatially established referents did positively influence reaction times. This finding is consistent with a theoretically motivated distinction between grammatical space and topographic use of space in sign. In sum,these studies have found that the same processing mechanisms are required to interpret co-reference in signed and spoken languages, but for signed languages, the type of information represented by the spatial location can influence how co-reference relations are processed (Emmorey, 2002).

H. Effects of Sign Language Structure on Memory

Classic theories of verbal working memory (Baddeley, 1986) propose that verbal information is maintained in memory via the existence and partial interaction of two components: a phonological buffer in which phonological representations of words are transiently stored, and an articulatory rehearsal mechanism through which those representations are continuously updated. This proposed architecture of verbal working memory accounts for four well-attested phenomena associated with spoken language immediate serial recall (discussed below)–the phonological similarity effect, the irrelevant stimulus effect, the word length effect, and the articulatory suppression effect (reviewed in Baddeley, 1990; Neath, Surprenant, & LeCompte, 1998; Wilson & Emmorey, 1997a).

However, it is not a priori obvious whether this architecture can account for working memory for sign language signs. While signs clearly have phonological structure that is amenable to verbal storage and rehearsal mechanisms, their visual-manual nature likely requires processing routines that draw heavily from those subs serving non-linguistic visual spatial working memory. In recent years, a great deal of knowledge about visual spatial working memory for sign language signs has come from the studies of Margaret Wilson and colleagues (e.g., Wilson & Emmorey, 1997a; Wilson, 2001a). These studies demonstrate that the similarities between spoken and sign working memory functions are profound, encompassing each of the phenomena on which classic working models are predicated.

One of the four effects that form the foundation of classic working memory models is that serial recall is poorer for words situated in the context of phonologically similar words (e.g., dog, log, lot, dot) than with words with different phonological representations. This phonological similarity effect (for a review see Gathercole & Baddeley, 1996) has been taken as evidence that the code by which words are stored in working memory is phonological in form. An analogous effect has been found in sign language serial recall, in which performance is worsened for signs that share phonological parameters, such as hand configuration, with other signs in a signed list (Hanson, 1982; Klima & Bellugi, 1979). That word similarity effects exist in signed languages is evidence that signs are stored in working memory by a phonological or kinesthetic code, just as spoken words. That this code is visual–manual rather than auditory–oral, however, is evidence that the entries in verbal working memory, no matter the architecture of the system, are themselves modality-specific representations.

Parallel patterns of serial recall disruption in speech and sign are also evidenced by performance decrements in the presence of irrelevant, modality-specific stimuli. Specifically, hearing subjects ’spoken word recall performance is known to be impaired when they are presented with irrelevant auditory stimuli, such as speech or non-linguistic tones, immediately following the presentation of a list of words to be remembered. This irrelevant sound effect (for a review see Neath, 2000) occurs when the words to be remembered are either auditory or printed. However, in contrast to findings from studies of the phonological similarity effect, effective interfering stimuli for speech recall are seldom or never visual (Wilson & Emmorey, 2003), presumably because the interfering visual stimuli used in irrelevant speech tasks are not coded into an auditory form capable of causing interference. In contrast, Wilson and Emmorey (2003) have found that irrelevant visual stimuli do reduce sign recall performance. Interestingly, both pseudo-signs and non-linguistic moving visual objects were markedly effective in doing so. This finding is consistent with a view of sign language working memory in which signs are stored as visual–manual phonological representations, and is further evidence of a common principle (irrelevant stimulus effects) being instantiated in a modality-specific way (auditory for speech, and visual for sign).

Both the phonological similarity effect and the irrelevant stimulus effect are associated with the buffer component of the classic verbal working memory model. In contrast, the rehearsal component is associated with two different effects, the word length effect and articulatory suppression. The word length effect is the finding that short words are remembered better than longer words on immediate serial recall tasks. Generally this is believed to be a result of the greater amount of time required to articulate (and thus rehearse) longer words, although it is controversial as to whether the critical factor is actually the number of syllables or phonemes in a given word, rather than the length of the word per se (Baddeley,Thomson, & Buchanan, as cited in Wilson, 2001a). By varying the physical distances that signs must traverse during articulation, Wilson and Emmorey (1998) were able to demonstrate that greater articulatory time, rather than a greater number of phonological units, is indeed responsible for the word length effect in sign languages. This study suggests that although spoken and signed languages call upon different articulators and motor programs in the service of word production, the processes associated with retaining words and signs in working memory are constrained by common properties.

Effects of articulation on memory span are not confined to word length. Span is also reduced when subjects engage in non-speech articulatory behaviors, while holding a list of words in memory. This finding, known as the articulatory suppression effect (Murray, 1968; Smith, Reisburg, & Wilson, 1992), has been shown by Wilson and Emmorey (1997b) to be robust for sign as well as speech. Specifically, signers' immediate serial recall performance was reduced when they opened and closed their hands (from an S hand to a 5 hand and back) repeatedly during stimulus presentation. The articulatory suppression effect is a testament to the importance of modality-specific articulatory rehearsal in verbal working memory. By engaging in an unrelated motoric task, planning and/or implementation of the motor program normally used to rehearse and produce a spoken or signed word is suppressed, and articulatory rehearsal is no longer possible. As a consequence, memory performance suffers.

Broadly, the functional architecture of working memory for spoken and signed language thus appear remarkably similar. Each has a storage buffer in which lexical items are represented by phonological codes – codes that are susceptible to interference both by the concurrent presence of highly similar phonological codes, and by the presence of irrelevant stimuli that share the modality – auditory or visual – of the words to be remembered. The lexical items in the phonological buffer, spoken or signed, decay unless sub-vocally or sub-manually rehearsed, a process that is itself sensitive to the length of the words to be remembered. Perhaps most interestingly, blocking articulatory rehearsal will not only result in worsened memory performance, it will suppress other effects that it gates, such as the phonological similarity effect (Baddeley, Lewis, & Vallar, 1984; Wilson & Emmorey, 1997a, 1997b). Articulation thus serves as a mechanism for getting words into the phonological buffer, and for keeping them active while there.

I. Neural Representation of Signed Languages

Our understanding of the neural representation of human language has been greatly enriched by the consideration of signed language of the deaf. Outwardly, this language form poses an interesting challenge for theories of cognitive and linguistic neural specialization, which classically has regarded the left hemisphere as being specialized for linguistic processing, while the right hemisphere is specialized for visual-spatial abilities. Given the importance of putatively visual-spatial properties of sign forms (e.g. movements trajectories and paths through 3-dimension space, memory for abstract spatial locations, assessments of location and orientation of the hands to the body ,etc.),one might expect a greater reliance of right-hemisphere resources during sign language processing. However, as discussed above, despite major differences in the modalities of expression, once we acknowledge the structural homologies of spoken and signed language forms, striking parallels in the psycholinguistic and cognitive processing of these languages emerge. Thus, these commonalities suggest a possible uniformity in neural systems, and the cognitive processes they mediate, underlying both signed and spoken processing.

Case studies of deaf signing individuals with acquired brain damage and neuroimaging studies of healthy deaf subjects have provided confirming evidence for the importance of left hemisphere system in the mediation of signed language. Deaf signers, like hearing speakers, exhibit language disturbances when left-hemisphere cortical regions are damaged (e.g., Hickok, Love-Geffen, & Klima 2002; Marshall, Atkinson, Smulovitch, Thacker, & Woll, 2004; Poizner, Klima, & Bellugi, 1987; for a review see Corina, 1998a, 1998b). In addition, there is good evidence that within the left hemisphere, cerebral organization in deaf signers follows the familiar anterior/posterior dichotomy for language production and comprehension, respectively, that is familiar from speech. In addition, neuroimaging studies have raised new questions regarding the unique role of the right hemisphere in sign language comprehension, as some evidence suggests that posteriorparietal regions may play a special role in the mediation of signed languages (Newman, Bavelier, Corina, Jezzard, & Weville, 2002).

J. Sign Language Aphasia

1. Sign Language Production

In spoken language aphasia, chronic language production impairments are typically associated with left-hemisphere frontal anterior lesions that involve the cortical zone encompassing the lower posterior portion of the left frontal lobe,e.g.,Broca’s area. These lesions often extend in depth to the periventricular white matter (e.g., Mohr et al., 1978; Goodglass, 1993). The anterior insula has also been implicated in chronic speech production problems (Dronkers, Redfren, & Knight, 2000). In the context of understanding prefrontal contributions to sign language, a pertinent example of left-hemisphere language mediation is that of patient G.D., reported in Poizner et al. (1987). G.D., a deaf signer with a large lesion in a left, anterior frontal region encompassing BA 44/45, presented with non-fluent, aphasic signing with intact sign comprehension. Specifically, G.D.’s signing was effortful and diffluent, with output often reduced to single-sign utterances. The signs she was able to produce were a grammatical, devoid of the movement modulations that signal morphed-syntactic contrasts in fluent signing. As with hearing Broca's aphasics, this signer's comprehension of others' language productions was undisturbed by her lesion. Both at the single sign and sentence level, her comprehension was on par with control subjects. That this deficit is not simply motoric in nature is indicated by the fact that the deficits were exhibited on both her meteorically and non-meteorically (i.e., ipsilesional) limb.

2. Sign Paraphasia

Sign language breakdown following left hemisphere damage is not haphazard, but affects independently motivated linguistic categories. This observation provides support for viewing aphasia as a unique and specific cognitive deficit rather than as a subtype of a more general motor or symbolic deficit. A fascinating example of the systematicity in sign and spoken language breakdown is illustrated through consideration of paraphasia errors (Corina, 2000). The substitution of an unexpected word for an intended target is known as verbal paraphasia. Most verbal paraphasias have a clear semantic relationship to the desired word and represent the same part of speech, hence, they are referred to as ‘‘semantic paraphasias’’(Goodglass,1993). In contrast, phonemic or ‘‘literal’’paraphasia refers to the production of unintended sounds or syllables in the utterance of a partially recognizable word (Blumstein, 1973; Good glass, 1993). Theoretically, sound distortions arising from phonetic impairment are not considered to be instances of paraphasia; however, in practice, it is quite difficult to distinguish true paraphasic errors from phonetic based sound distortions. Phonemic sound substitution may result in another real word, related in sound but not in meaning (e.g., telephone becomes television). Also attested are cases in which the erroneous word shares both sound characteristics and meaning with the target (broom becomes brush; Good glass, 1993).

Several reports of signing paraphasia can be found in the sign aphasia literature. In an early report of ‘‘neologistic’’signing, Leischner (1943) describes a deaf subject with left hemisphere damage who produced ‘‘fluent but nonsensical signing.’’ Unfortunately, little description of these errors was provided. Several well-documented examples of semantic paraphasias have been reported (Poizner et al., 1987; Brentari, Poizner, & Kegl, 1995; Corina et al., 1992). For example, subject P.D. (Poizner et al., 1987) produced clear lexical substitutions: BED for CHAIR, DAUGHTER for SON, QUIT for DEPART,etc. In general,the semantic errors of P.D. overlap in meaning and lexical class with the intended targets; this pattern has been routinely observed in spoken language semantic paraphasia. Subject W.L. (Corina et al., 1992) evidenced interesting semantic blends in signing, errors conditioned, in part, by perseverations from earlier cued items. For example, in the context of a picture-naming task, when shown a picture of a tree, W.L. signed TREE with the G hand shape. Previously, W.L. had been asked to name the color green. The lexical signs GREEN and TREE share a motion (twisting of the wrist) and evidence similar articulatory postures. These ASL semantic paraphasias suggest that the lexicon is structured according to semantic principles, whereby similar semantic items share representational proximity. In this view, co-activation of closely related representations and/or an absence of appropriate inhibition from competing entries may lead to substitutions and blends.

One of the most striking characteristics of aphasic signing is formational paraphasia. As with spoken languages, ASL formational errors encompass both phonological and phonetic levels of impairment (see Corina, 2000, for some discussion). A priori, we may expect to find true phonological errors affecting the four major formational parameters of ASL phonology: hand shape, movement, location, and orientation. However, the distribution of paraphasic errors among the four parameters of sign formation appears to be unequal; hand shape configuration errors are the most widely reported, while paraphasias affecting movement, location, and orientation are infrequent (see, e.g., Poizner et al., 1987). Not only have hand shape errors been observed across many different aphasic signers, but the frequency of occurrence in individuals who exhibit this disruption is quite high. The globally aphasic signer W.L. (Corina et al., 1992) produced numerous phonemic errors, nearly all of which were errors involving hand shape specification. For example,W.L. produced the sign TOOTHBRUSH with the Y hand shape rather than the required G hand shape, and produced the sign SCREWDRIVER with an A hand shape rather than the required H hand shape. The higher incidence of hand shape errors is interesting, as recent linguistic analyses of ASL have suggested that hand shape specifications (and perhaps static articulatory locations) may be more consonantal in nature, while movement components of ASL may be analogous to vowels (see Corina & Sandler, 1993, for some discussion). In spoken language phonemic paraphasias, a homologous asymmetry exists; the vast majority of phonemic paraphasias involve consonant distortions. Another similarity between spoken and sign paraphasic error is that in each case, errors do not compromise the syllabic integrity of a sign or word (Brentari et al., 1995; Corina, 2000).

3. Sign Language Comprehension

Fluent spoken language aphasias are associated with lesions to left-hemisphere posterior temporal regions. Wernicke's aphasia, for example, is often associated with damage to the posterior regions of the left superior temporal gyros. More recent work has suggested the contribution of posterior middle temporal gyrus in cases of chronic Wernicke's aphasia (Dronkers, Redfern, & Ludy, 1995; Dronkers et al., 2000). Two prominent features of Wernicke's aphasia are impaired comprehension and fluent, but often paraphasic (semantic and phonemic) output. Additionally, persistent neologistic output sometimes occurring with severe Wernicke's aphasia is associated with lesions extending to the supra marginal gyrus (Kertesz, 1993).

Signers with left-hemisphere posterior lesions also evidence fluent sign aphasia. Two cases are reported in Chiarello, Knight, and Mandell (1982), Poizner et al. (1987), and Corina et al. (1992). These patients presented with severe comprehension difficulties in the face of relatively fluent but paraphasic output. Interestingly, while damage to left temporal cortex has been demonstrated to impair sign comprehension in some patients (Hickok et al., 2002), the lesions in the two case study patients above did not occur in cortical Wernicke's area proper, but rather involved more frontal and inferior parietal areas. In both cases, lesions extended posteriorly to the supra marginal gyrus. This is interesting, as lesions associated with the supra marginal gyrus alone in users of spoken language do not typically result in severe speech comprehension deficits. These two cases suggest that sign language comprehension may be more dependent than speech on left hemisphere inferior parietal areas, a difference that may reflect within-hemisphere reorganization for cortical areas involved in sign comprehension (Leischner, 1943; Chiarello et al., 1982; Poizner et al., 1987).

4. Cortical Stimulation Mapping

Additional insights into the neural control of paraphasic errors have been reported by Corina et al. (1999), who investigated sign language production in a deaf individual undergoing an awake CSM procedure for the surgical treatment of epilepsy. During the language mapping portion of the procedure, a subject is required to name pictures or read written words. Disruption of the ability to perform the task during stimulation is taken as evidence of cortical regions integral to the language task (Stemmer & Whitaker, 1998).

In this deaf patient, all testing was conducted using ASL. The patient was to sign the names of line drawings of pictures. All signs were one-handed, and the subject signed each with his left hand. Because this subject was undergoing left-hemisphere surgery, language disruption as a result of cortical stimulation cannot be attributed to the suspension of primary motor functioning.

Stimulation to two anatomical sites led to consistent naming disruption. One of these sites, an isolated frontal opercula site, corresponds to the posterior aspect of Broca's area, BA 44. A second site, located in the parietal opercula region, also resulted in robust object-naming errors. This parietal area corresponds to the supramarginal gyrus (SMG, BA 40). Importantly, the nature of these errors was qualitatively different. Stimulation of Broca's area resulted in errors involving the motor execution of signs. These errors are characterized by a laxed articulation of the intended sign, with nonspecific movements (repeated tapping or rubbing) and a reduction in handshape configurations to a laxed-closed fist handshape. Interestingly, there was no effort on the part of S.T. to self-correct these imperfect forms. Our results are consistent with the characterization of the posterior portion of Broca's area as participating in the motoric execution of complex articulatory forms, especially those underlying the phonetic level of language structure.

The sign errors observed with stimulation of the SMG are qualitatively different. With stimulation to this site, S.T. produced both formational and semantic errors. Formational errors are characterized by repeated attempts to distinctly articulate the intended targets, commonly with successive formational approximations of the correct sign. For example, the sign PEANUT is normally signed with a closed fist and outstretched thumb with a movement composed of an outward wrist rotation (the thumb flicking off the front of the teeth). Under stimulation, this sign began as an incorrect but clearly articulated, ‘‘X’’handshape (closed fist with a protruding bent index finger) produced at the correct location, but with an incorrect inward rotation movement. In two successive attempts to correct this error, the subject first corrected the handshape, and then went on to correct the movement as well. Notably, we do not find the laxed and reduced articulations characteristic of signing under conditions of stimulation to Broca's area. Instead, as these examples illustrate, under stimulation to the SMG, the subject's signing exhibits problems involving the selection of the individual components of sign forms (i.e., handshape, movement, and, to a lesser extent, location).

Semantic errors were also observed under stimulation of the SMG, and the characteristics of these errors are particularly noteworthy. Specifically, all of these errors involve semantic substitutions that are formational quite similar to the intended targets. For example, the stimulus picture ''pig' 'elicited the sign FARM; the stimulus picture ''bed'' was signed as SLEEP; and the stimulus picture ''horse'' was signed as COW. In ASL, these semantic errors contain considerable formational overlap with their intended targets. For example, the signs PIG and FARM differ in movement, but share an identical articulatory location (the chin). Each is made with a similar hand shape; the signs BED and SLEEP share hand shape and are both articulated about the face; finally, the signs COW and HORSE differ only in hand shape. In English, these mistakes might be similar to uttering "lobster" when one intended to say "oyster," or "plane" when one intended to say "train". That is, these errors share both semantic and formational properties.

In summary, the analysis of these data suggests that stimulation to Broca's area has a global effect on the motor output of signing, whereas stimulation to parietal opercula site (the SMG) disrupts the correct selection of the linguistic components (including both phonological and semantic elements) required in the service of naming.

K. Neuroimaging Studies

Neuroimaging techniques like PET and fMRI also make unique contributions to our current understanding of the neurological processing of signs. In particular, these studies reaffirm the importance of left-hemisphere anterior and posterior brain regions for sign language use and emphasize that some neural areas appear to participate in language perception and production, regardless of the modality of the language.

Sign language production tasks are especially likely to recruit the left hemisphere. For example, when signers name objects (Emmorey et al., 2003), generate verbs to accompany nouns (e.g., CHAIR → SIT) (McGuire et al., 1997; Petitto, Zatorre, Gauna, Nikelski, Dostie, & Evans, 2000; Corina, San Jose-Robertson, Guillemin, High, & Braun, 2003), or sign whole sentences (Braun,Guillemin,Hosey,& Varga,2001),their left hemispheres show significant increases in blood flow, relative to control tasks. It has been suggested that this heightened blood flow reflects, in part, the activation of motor systems needed for the production of complex linguistic actions.

Sign language comprehension also recruits the left hemisphere in some studies, for both word- and sentence-level tasks. For example, classic Broca's area has been found to be involved in sign comprehension when subjects observe single signs (Levanen, Uutela, Salenius, & Hari, 2001; Petitto et al., 2000) and sentences (Neville et al., 1998; MacSweeney et al., 2002). This activation is not limited to anterior regions. When signers of BSL view their language, posterior left-hemisphere regions are activated, including the posterior superior temporal gyrus and sulcus, and the supramarginal gyrus (MacSweeney et al., 2004). This heightened activation is relative to complex nonlinguistic gestures, and does not occur for non-signers.

Interestingly, there is growing evidence that right-hemisphere regions may also be recruited for aspects of sign-language processing in ways that are not required in the processing of spoken languages. At least one sign-language production task is known to recruit right-hemisphere brain regions. When deaf signers were asked to use classifier constructions to describe the relative positions of two objects depicted in a line drawing, both left- and right-hemisphere regions were found to be active (Emmorey et al., 2002). When ASL prepositions were used instead of classifiers, only the right hemisphere was recruited.

Other evidence suggests that right-hemisphere posterior parietal regions may contribute to the processing of some aspects of sign comprehension (Bavelier et al., 1998; Capek et al., 2004; Corina, 1998b; Newman et al., 2002). For instance, both left- and right-hemisphere cortical regions were recruited when hearing native signers of ASL passively watched ASL sentences (Newman et al., 2002). Some right-hemisphere structures appear to be specialized for processing spatial information, including biological motion.

It may be that ASL phonological distinctions that make use of space are the trigger for right-hemisphere recruitment in sign perception.

Moreover, right-hemisphere involvement may be related to the age at which the signer first acquired a sign language. One particular structure, the right angular gyrus, was found to be active only when hearing native users of ASL performed the task. When hearing signers who learned to sign after puberty performed the same task, the right angular gyrus failed to be recruited. Thus, the activation of this neural structure during sign language perception may be a neural “signature” of sign competence developing during the critical period for language (Newman et al., 2002).

In sum, while left-hemisphere regions are undisputedly recruited in a similar fashion for both sign and speech, it has been argued that the right-hemisphere activation seen during sign language comprehension is more robust than that observed in studies of spoken language processing. Continued research using cognitive neuroscience tools such as PET and fMRI will provide more opportunities to investigate these findings.

In conclusion, preliminary findings from psycholinguistic and cognitive studies of signed languages indicate a great deal of commonality between the recognition, access, and memory structures of the representations of spoken and signed words in the mental lexicon. These similarities extend to those processes underlying the parsing and interpretation of morphological and syntactic structure. Moreover, burgeoning cognitive neuroscience research has also begun to specify the commonalities and differences between neural systems underlying spoken and signed language forms. While some unique and subtle differences have been found between their respective processing, overwhelming evidence suggest a great deal of homology between speech and sign neural representations. These data suggests that core properties underlie the neural capacity for human language, regardless of the surface form taken by the linguistic communicative system. Future studies will doubtless continue to further specify both the common and unique aspects of these forms of language while striving to better understand the processing domains under which the modality of language expression does and does not affect the final form of the mental structure of language. In this way, we may come to understand both the biological and environmental contributions that shape human linguistic communication.

CHAPTER 15
LANGUAGE LEARNING IN INFANCY

INTRODUCTION

All humans have equal rights, just like children who have the right to grow and develop, play, rest, play, and learn in an education. So, learning is a child's right not an obligation. Parents and the government must provide educational facilities and infrastructure for children in the context of learning programs. Because learning is a child's right, learning must be fun, conducive, and allow children to be motivated and enthusiastic. Obtain stimulation of basic abilities to the development of language, cognitive, physical motor and art, as well as the development of habits that consist of religious values, social, emotional and independence., The basic abilities of children support each other. children namely language skills. Every human has the same potential to master the language. The process and nature of each person's language acquisition takes place dynamically and through stages. Humans begin their communication with the world around them through crying. A baby trains the language by communicating all his needs and desires. In line with the development of physical abilities and maturity, especially those related to the process of speech, communication is increasingly increasing and expanding, for example, with those around him, the environment and developing with other people who are newly known and friendly with him. The language development is always increasing in accordance with the increasing age of the child. Language development in children is very important because children can develop social skills (social skills) through language. Through language, children can express their thoughts using language so that others can capture what is thought by children and create a social relationship. In time the child will be able to develop and grow into a happy person because by starting to communicate with the environment, willing to give and accept everything that happens in the environment. The development process goes through various stages of the child's language development, from childhood to school-age mastery. In this stage of language acquisition, the role of parents as the closest people is needed. Parents should always pay attention to these developments, because at this time it really determines the process of a child in socializing and learning.

A. Language Learning in Infancy

Learning language is one of the most impressive and intriguing human accomplish- ments. Think about the vast differences between the healthy 12-month-old child who says “Ah! Ah!” with hands held up in the air, eager to be lifted from the highchair, and the same child six months later using recognizable two-word combinations coordinated with gestures (e.g., “Mommy out!”). Within the next year that child will start using an impressive complement of morphosyntactic skills to produce utterances that reflect con- siderable linguistic sophistication (e.g., “Mom! I wanna get outta this chair now!”). The child’s desire may be equally intense in each of these situations, yet clearly the typical 2-year-old has advanced significantly beyond the 1-year-old in her ability to effectively use her native language to make that particular desire known to those around her. The goal of developmental psycholinguistics is to map the endogenous and exogenous forces that converge to shape and guide this set of developmental achievements. Over the past five decades, the field of language development research has been at the center of the debate between nativist and constructivist approaches to understanding human cognition. In the early 1960s, Chomsky’s proposal that language acquisition was innately guided by a Language Acquisition Device offered a powerful solution to the logical problem of how children learn language, a view still ardently embraced by many in the field. Since that time, however, an alternative view has been gaining momentum, gathering logical and empirical support for the idea that a child’s linguistic knowledge is constructed rather than triggered, emerging as a consequence of the child’s experi- ences with the linguistic and non-linguistic world (e.g., Bates & MacWhinney, 1979; Braine, 1976; Slobin, 1973). Proposals on how exactly the child accomplishes this task have taken several different forms over the years, and with each new decade, theoretical and methodological advances have strengthened the case for this alternative to nativism. The goal of this chapter is to outline some key features of current proposals on how the child constructs a language. We first briefly review the standard nativist approach, and then discuss some recent developments in theory and research from diverse disciplines that have contributed to a shift in emphasis in research on language development in pro- ductive new directions. Then we review three lively areas of current research on lan- guage learning in infancy: early speech perception, lexical development, and listening for meaning. This review will of necessity be quite eclectic, focusing on a few studies within each of these areas that exemplify new perspectives that are now coming to the forefront in this field.

B. Nativist Views of Language Development

How and why does the child's linguistic behavior change so dramatically over such a short period of time? Much of the research on this topic has assumed that this process s driven by innate and highly specialized mental structures (a "mental organ," Chomsky, 1981). That is, learning language involves the operation of a specifically linguistic matu- rational bioprogram (i.e., Universal Grammar) as it processes specifically linguistic input (e.g., Chomsky, 1981; Lenneberg, 1967). According to this perspective, this innately spec- ified system is what makes it possible for the child to determine which of all the possible linguistic rule systems characterizes their particular native language. Indeed, the goal of much research in modern linguistics has been to map the diverse set of principles and fea- tures that describe the rule systems of any and all of the world's languages. Of course, in setting out on this daunting task, one is soon struck by the vast richness and complexity of human grammars. If the central question is how grammars come to be mastered, such complexity is par- ticularly disheartening, especially in light of the prevailing assumption among nativist theorists that the environment falls far short of providing what children need in order to learn rich systems of linguistic representations on their own. According to Chomsky (1981), speech by adults is so full of hesitations, false-starts, mispronunciations, and ungrammaticalities that it could not possibly be an adequate model from which to abstract complex and subtle linguistic regularities. Even if one acknowledged that child- directed speech is typically more coherent in structure than adult conversation (e.g., Snow & Ferguson, 1977), ethnographic research reveals substantial differences in the extent and nature of linguistic interactions with infants across cultures (e.g., Schieffelin & Ochs, 1986). Thus, it was apparently impossible to identify a universal set of features of child-directed input necessary for acquisition to take place (e.g., Lieven, 1994). Moreover, the wisdom of relying on a simplified child-directed register as the basis for grammatical development was called into question. Since speech to children is typically simpler than speech to adults, learning a grammar may be hindered by the fact that the input is limited in the scope and extent of the detailed syntactic information it can provide (Gleitman et al, 1988). Most significantly, several studies have demonstrated that caregivers do not provide enough explicit information to prevent the child from building overly general grammatical systems. This mistake, it was assumed, can only be overcome by linguistic input that provides “negative evidence”, i.e., information about what sentences are not permitted by the target language (Marcus, 1993, cf. Sokolov & Snow, 1994). These limitations of the input must be interpreted in the context of assumptions regarding what types of mechanisms are used by the language-learning child. In a famous examination of what would make languages “learnable,” Gold (1967) proposed that children are general learners who test hypotheses about grammatical rules against example sentences that they hear in the target language. In this demonstration, Gold provided what was considered to be compelling evidence that a general learner cannot induce the gram- mars of certain types of formal languages (which are derived from the class of natural languages) if it receives only “positive evidence” (i.e., only sentences that are gramma- tical in the language). The only conditions under which learning could be successful are: (1) if the learner is provided with negative evidence (i.e., cues to what sentences were ungrammatical), or (2) if the learner possessed strong initial biases about the types of hypotheses to consider in the first place. We should note that Gold’s learner incorporated an all-purpose learning mechanism that is clearly unlike any that has been proposed for young children. However, since several studies showed that explicit negative evidence is rare and is not universal in child-directed speech (e.g., Brown & Hanlon, 1970; Marcus, 1993), the only conclusion deemed reasonable at the time was that children must come pre-wired with a universal set of representational constraints on the types of grammars that are possible in human languages (e.g., Pinker, 1979). In other words, the complexity of the end-product and the indeterminacy of input to children (i.e., the “poverty of the stimulus”) appear to comprise compelling evidence that grammars cannot be learned. As Tomasello (2003) has recently put it, this view assumes that adult grammars go far beyond what children are capable of building given the resources available to them, i.e., “you can’t get there from here” (p. 2). Since most children do become relatively profi- cient at grammar within the first few years of life, it was logical to assume that it was only via a rich system of innately specified rules and representations, i.e., Universal Grammar, that all children could possibly zero-in on the particular set of rules that cha- racterize their native language (e.g., Chomsky, 1975; Pinker, 1999, 2003). Finally, this view also makes strong claims regarding the relations between linguistic and non-linguistic cognition. That is, the language faculty involves special processing mechanisms that are specifically dedicated to mediating the acquisition and processing of language. Moreover, sub-systems are themselves “modularized” in terms of components of the language faculty as traditionally defined by linguists, i.e., phonology, semantics, grammar, pragmatics (e.g., Fodor, 1983; Pinker, 1991; Levelt, 1989). Because these sub- components of language are assumed to be distinct in terms of the representations they employ, they are viewed as structurally autonomous and informationally encapsulated, not only from each other but also from the rest of non-linguistic cognition (e.g., Pinker, 1991). In sum, this nativist view of language development focuses on the specificity of the young child’s complex grammatical knowledge, the biological origins of its nature, and the uni- versal course of its acquisition. As framed by Pinker (1994): Language is a complex, specialized skill, which develops in the child sponta- neously, without conscious effort or formal instruction, is deployed without aware- ness of its underlying logic, is qualitatively the same in every individual, and is distinct from more general abilities to process information or behave intelligently. For these reasons, some cognitive scientists have described language as a psycho- logical faculty, a mental organ, a neural system, and a computational module. But I prefer the admittedly quaint term ‘instinct’. It conveys the idea that people know how to talk in more or less the sense that spiders know how to spin webs...spiders spin spider webs because they have spider brains, which give them the urge to spin and the competence to succeed (p. 18).

C. Learning

Although nativist views of language acquisition are forceful and still widely endorsed, there has been ongoing controversy about the adequacy of such theories as an account of how children develop competence in language. Some critiques directly challenge the logic of arguments made by Chomsky, Pinker, and like-minded theorists, questioning such core assumptions as the universality of generative grammar, the autonomy of syn- tax in language processing, and the fundamental unlearnability of language (e.g., Bates & Goodman, 1999; Braine, 1994; Pullum & Scholz, 2002; Tomasello, 1995).

Other critiques focus on empirical evidence inconsistent with particular nativist assertions. For example, the claim that negative evidence is not available when children make gramma- tical errors, an assumption central to the "poverty of the stimulus" argument at the heart of Chomsky's theory, is not supported by a recent analysis of parents' reformulations in speech to children (Chouinard & Clark, 2003). These diverse challenges, both philo- sophical and data-driven, have fueled debate over four decades about the explanatory adequacy of nativist theories of language learning. However, in recent years this debate has begun to change in focus and tenor, not only in response to explicit critiques within linguistics and developmental psychology, but also in response to research findings and theoretical insights from farther afield. An alternative perspective on language learning has been gathering force, amplified by new developments in research areas that formerly made little contact with theoretical debates on the nature of language development (see Kuhl, 2004; Seidenberg & MacDonald, 1999; Tomasello, 2003). We focus on four such developments that have begun to change the direction of research on early language acquisition: first, the emergence of more "user-friendly" theories of language and language use; second, the contribution of com- putational approaches to modeling language processing and learning; third, provocative findings from experimental research on learning and cognitive processing by infants; and fourth, insights from studies with children and non-human primates on the role of social cognition in communication. In different ways, these diverse areas all motivate and support an emerging alternative view of language learning.

1. New Ways of Understanding Language and Language Use

While generative theories have favored a view of linguistic competence defined exclusively in terms of grammatical knowledge, recent developments in both linguistics and psycholinguistics have shifted the focus to a more inclusive view of competence, one that incorporates performance factors that guide language use. Within the newly emerging area of cognitive-functional linguistics, "usage-based" theorists emphasize the essential connection between the structure of language and how language is used to communicate (e.g., Croft, 2001; Goldberg, 1995). According to this view, linguistic competence cannot be reduced to knowledge of a core grammar as Chomsky claimed, but rather draws on a wide range of cognitive and social capabilities and on knowledge from diverse domains. Rejecting fundamental nativist assumptions about the nature of language, usage-based theories demand a very different view of what is involved in language learning (Tomasello, 2003). In psycholinguistics as well there has been a dramatic shift away from models of speech processing that embody nativist assumptions, in favor of models that emphasize statistical and probabilistic aspects of language (Seidenberg, 1997). Until recently, the dominant processing theories have been those that presuppose the modularity of lan- guage, focusing on syntactic parsing strategies presumed to be automatic (see Frazier, 1987). For example, adults reading potentially ambiguous sentences such as “Put the apple on the towel into the box” are confused in predictable ways, presumed to result from of an irresistible initial tendency to interpret the prepositional phrase (PP) “on the towel” as modifying the verb and thus specifying the destination of the action. In fact, the first PP modifies “apple” rather than “put,” a reduced relative clause that attaches to the noun phrase rather than the verb phrase, catching the reader by surprise. Such clas- sic “garden-path” effects were replicated in hundreds of experiments based on nativist assumptions, providing support for the idea that default syntactic processing strategies are automatic and impervious to influence from other sources of information. But this picture is changing, with the emergence of new experimental paradigms that use more revealing techniques for monitoring on-line comprehension. For example, when adults are able to look at a relevant visual scene while hearing “Put the apple on the towel into the box,” different results emerge (Tanenhaus, Spirey-Knowlton, Eberhard & Sediry, 1995). If a single apple on a towel is present in the scene along with a second towel and a box, listeners look briefly at the lone towel when they hear the first PP, before look- ing at the box. This is the behavioral equivalent of the garden-path effect observed pre- viously in reading studies. However, if two apples are present in the scene, one on a towel and the other on its own, the same sentence is no longer perceived as ambiguous and there is no evidence of misinterpretation. To the contrary, the presence of the sec- ond apple provides immediate non-linguistic contextual support for interpreting “on the towel” as modifying the noun, crucial information that enables the listener to iden- tify the correct referent. This experiment by Tanenhaus et al. was one of many recent studies using eye-tracking techniques to show that listeners integrate probabilistic in- formation from multiple sources in interpreting spoken language, rather than defaulting to inflexible syntactic processing strategies (see Trueswell & Tanenhaus, 2005). These new theoretical perspectives on language use emerging within linguistics and psycholinguistics are not only “user-friendly” in their emphasis on the flexibility and resourcefulness of mature language processing, but also “child-friendly” in their developmental implications. Their influence on current research on early language learn- ing is evident at several levels. Developmental theorists using observational techniques are testing predictions from usage-based theories, namely that children develop linguis- tic competence gradually, learning to produce new constructions item by item, rather than advancing by triggering innately specified grammatical rules that function in an all- or-none fashion (e.g., Lieven, Pine, & Baldwin, 1997; Tomasello, 2003). Researchers are also using new experimental techniques for assessing early speech-processing abilities to explore how infants in the first year track distributional information in spoken language (Saffran, Werker, & Werner, in press), and how by the second year they are able to inter- pret speech incrementally based on probabilistic information, similar to adults (Fernald, McRoberts, & Swingley, 2001). These are just a few examples to illustrate the more gen- eral point, that the application of probabilistic constraints is not only central to under- standing adult competence in language processing but is now being extended to theories of language acquisition as well. Seidenberg and MacDonald (1999) make the case that the processes of constraint satisfaction that are critical to mature language production and interpretation are the same processes used by infants as they begin to make sense of speech and break into language.

2. Computational Approaches to Language Use and Language Learning

The idea that attending to distributional information in speech might be critical in lan- guage learning was proposed many years ago (Maratsos & Chalkley, 1980), at a time when nativist views of language dominated the field. Another reason why this early interest in distributional learning was initially eclipsed was that the computational resources necessary for exploring such questions empirically were not yet widely avai- lable. In recent years, however, computational approaches of different kinds have become increasingly influential in research on language development, ranging from statistical analyses of language patterns to connectionist models. Because large corpora of sponta- neous speech by parents and children are now accessible through the Child Language Data Exchange System (CHILDES) data bank (MacWhinney & Snow, 1990), researchers are able to undertake detailed analyses of the kinds of distributional infor- mation available in the language directed to the child. For example, statistical models have been used to reveal cues in child-directed speech that could potentially aid the young language learner in identifying word boundaries (e.g., Brent & Cartwright, 1996; Christiansen, Allen, & Seidenberg, 1998; Swingley, 2005) and in classifying new words in the appropriate grammatical form class (e.g., Mintz, 2003). Statistical models such as these provide evidence that information about the distribu- tion of linguistic units at various levels is available in the speech stream that could, in principle, facilitate learning by the child. Connectionist models are well suited to tackle the next question, asking what kinds of outcomes are possible at different phases in development given a particular input and a general-purpose learning mechanism (Elman et al., 1996). These models typically represent information in a distributed fashion across a set of connections between input and output units, although representational features and network architectures have been varied in many interesting ways (e.g., Shultz, 2003; Munakata & McClelland, 2003). Over training, the networks extract and represent patterns of regularities in the input, abstracting information from multiple sources simul- taneously and at multiple levels of granularity. Guided by the non-linear learning mech- anism, the networks allow solutions to be represented as the coordinated activity of the network as a whole (i.e., the patterns of connections across the weights) that can be eval- uated at different points in the training. For example, in a series of models of the acquisition of inflectional morphology (Plunkett & Marchman, 1991), networks were trained to map words from their stem (e.g., walk) to past tense (e.g., walked) forms using artificial languages with different propor- tions of regular and irregular verbs. These factors were evaluated in a parametric fashion, revealing information about the conditions under which the networks would make use of phonological regularities across stem–past tense pairs as well as the role of both token frequency (i.e., how many times a particular stem–past tense mapping was seen) and type frequency (i.e., how many different stem–past tense pairs shared the same overall pattern, e.g., ring-rang, sing-sang) in determining learning. These features predicted the learning patterns in the networks and have been examined in several studies of natural languages (e.g., Bybee, 1995). Plunkett and Marchman (1993) next examined learning of stem–past tense mappings in the context of a lexicon that gradually increased in size over the course of the training, i.e., incremental learning. That is, the ability of the networks to memorize particular stem–past tense pairs or to generalize to novel forms was sensitive to develop- mental changes in the overall size and composition of the training set. Again, the predic- tions of these models have led to examinations of the role of vocabulary size in children's learning of morphosyntax and the impact of individual differences in vocabulary size on later grammatical outcomes (e.g., Marchman & Bates, 1994; Bates & Goodman, 1999). Another example is Elman's (1993) simple recurrent network (SRN) model which ab- stracted syntactic regularities across sequential occurrences of lexical items. The model was presented strings of words in "sentences" generated by a human-like artificial gram- mar. The task for the SRN was to predict the upcoming word in the sequence, a task that is inherently probabilistic given the many possible words that could come next, especially across sentence boundaries. However, the task of "listening ahead" in essence forced the network to track distributional relations across the words, encoding the syntax of the sentences in terms of the varying conditional probabilities that were inherent in the example sentences. Interestingly, the network was successful in this task only when limitations were place on the size of the "working memory" early in learning and then memory size was gradually increased across the course of training. The network's limi- tation, as it turned out, was an advantage in learning the syntax of this artificial language, illustrating the importance of "starting small" (see also Newport, Bavelier, & Neville, 2001; Hertwig & Todd, 2003). While "starting small" may not always be necessary for successful learning (Rohde & Plant, 1999) and there are clearly limitations to what these models can tell us about human learning, such endeavors have served as natural tools for testing the conditions under which knowledge can emerge with exposure to different kinds of learning environments (e.g., Kersten & Earles, 2001). In addition, they have fueled new interest in exploring the ways in which knowledge is used and represented, what kinds of information are available to the language-learning child, and what types of learning mechanisms can account for both the general patterns and individual variation seen across development (Elman et al., 1996; Elman, 2004, 2005).

3. Learning Strategies in Infancy

A third area in which recent developments have challenged traditional assumptions about language acquisition is research on infant learning strategies. When Skinner (1957) proposed that language was a behavior like any other animal behavior that could only be learned through gradual shaping and external reinforcement, Chomsky (1959) emphati- cally rejected this idea. The common observation that children begin to talk correctly in the absence of explicit positive or negative feedback was clearly at odds with the idea that language learning is achieved through a process of conditioning. Chomsky's claim that principles of learning could not possibly explain how children master language was based on the behaviorist learning theory of the day, which was much too simplistic to account for the complexities of children's linguistic ability. In recent years, however, researchers investigating early perceptual and cognitive development have made stunning discoveries about the learning capacities of young infants, and a new view of the poten- tial role of learning in language development has emerged. Thus, while computational models reveal how much information about language structure is potentially available to the young language learner, these new behavioral studies confirm that infants are in fact able to learn from this information.

For example, Saffran, Newport, and Aslin, (1996) showed that eight-month-olds can segment a stream of meaningless syllables containing no acoustic or prosodic cues to word boundaries after only a few minutes of listening experience. The information infants are using to identify word-like units in this case is distributional evidence, the regularities in the relative position and order of particular syllables over the whole sequence. For exam- ple, one string of syllables consisted of pa bi ku go la tu da ro pi ti pu do da ro pi go la tu... After familiarization with this sequence, infants were tested with "words" that had occurred in the string, i.e., sequences of syllables that always occurred in the same order, such as pabiku and golatu. They were also tested with "non-words," combinations of fa- miliar syllables that spanned two different words, such as kugola. Although there was no acoustic information specifically marking word boundaries, the transitional probabilities were much higher between syllables within words than between words. Thus there was statistical information that could enable infants to identify the familiar word-like units in the stream of speech. The finding that they are capable of performing such computations reveals the sophisticated talents young learners bring to the task of segmenting speech, months before they are able to understand meanings in the words they hear.

4. Social Cognition in Infants and Non-Human Primates

Another domain of research that is yielding surprising findings relevant to language learning focuses on the abilities of human infants and animals of other species to appreciate the mental states of others. Ethologists have long been interested in the com- municative function of animal signals, attempting to establish criteria for determining whether primate vocalizations constitute intentional and semantically meaningful signals to conspecifics. Vervet monkeys, for example, give at least three acoustically distinctive calls in response to particular predators (Cheney & Seyfarth, 1997). When one of these call types is played back in field experiments in the absence of an actual predator, adult vervets react appropriately, taking a different escape route in response to a “snake call” than to an “eagle call.” These and other recent ethological observations have dispelled the prevailing assumption that primate communication is entirely reflexive and emotional in nature and thus must be very different from human communication. These findings are of interest to many outside the field of biology, posing questions of concern to philo- sophers, linguists, and psychologists. What sorts of mental representations underlie the production and perception of these primate calls? When a vervet gives a snake alarm call, is this analogous to a child crying out “Snake! Watch out!” on seeing slithery movement in the grass? Cheyney and Seyfarth (1997) conclude that vervet and human communication differ fundamentally, in that monkeys call and look at each other in order to influence each other’s behavior, whereas children do so in order to influence their attention or know- ledge. According to these researchers, the lack of a theory of mind in vervets is one fun- damental reason they are incapable of language. However, these questions are complex, and studies with other species show that animals as diverse as sea lions, parrots, and bonobos are capable of learning to use symbols and to distinguish between objects, actions, and modifiers. There is energetic debate about how these abilities should be interpreted and how they differ among species, in particular whether chimpanzees and other higher primates show more sophisticated competence in reading the goals and intentions of others than do monkeys (Tomasello & Carpenter, 2005). This new wave of research on animal communication resonates in interesting ways with two flourishing areas of research with human children. The first area explores the growth of theory of mind in preschoolers (e.g., Wellman, 2002), while the second focuses on children's early sensitivity to communicative behaviors such as emotional expres- sions, gaze direction, and pointing (Baldwin & Moses, 1996). Experimental studies using looking-time measures show not only that young infants can use such vocal and visual cues to guide their attention (e.g., Mumme & Fernald, 2003), but also that they use non-verbal referential cues in combination to make inferences about the goals and intentions of others (see Rochat, 1999). It may not be obvious what these studies of non-verbal communication in animals and infants have in common with the paradigms described in the three previous sections, which all seem more directly relevant to the new focus in language research on how children and adults evaluate multiple sources of probabilistic information in interpreting and using language. But the relevance of research on inten-tionality to this new focus will become clear when we review current research on lexical development. Through the work of Tomasello (2003) and others, it is increasingly apparent that the ability to learn what a word means and to use words in communication depend crucially on fundamental skills of joint attention and intention-reading. Because language use involves mind-reading, children must learn to interpret others' mental states and to integrate probabilistic information on this level with information at other linguis- tic and non-linguistic levels in order to interpret language and to communicate effectively.

In summary, for half a century research on language development has been domi- nated by theoretical claims and assumptions emerging from a powerful linguistic the- ory that defines competence in terms of innately specified grammatical knowledge. Recent developments in diverse fields outside the mainstream of linguistics are forcing re-examination of these assumptions from different angles, all suggesting alternative ways of thinking about how human language functions and what it means to learn a language. A dominant theme emerging from the new paradigms and findings in the four areas reviewed above is that human communication relies on the integration of many different sources of information, rather than on innately specified knowledge in an encapsulated, modular system isolated from other cognitive and social capacities. Seidenberg and MacDonald (1999) refer to this emerging perspective as the "proba- bilistic constraints" approach, which has as its central idea that language learning by children and language processing by adults both involve the use of “multiple, simulta- neous, probabilistic constraints defined over different types of linguistic and nonlin- guistic information” (p. 570). This theme is reflected in some of the most interesting new research on infant speech perception, early word learning, and the emergence of efficiency in spoken language un- derstanding, to be reviewed in the following sections. The purpose of this brief review is to point to a few of the recent contributions to research on language acquisition in infancy that exemplify this sort of probabilistic perspective. We do not attempt to provide a com- prehensive review of this area and certainly do not assume that this approach has all the answers. Instead, our goal is to describe new research on how children start out by attending to patterns in speech sounds in the first year and learn to listen for meaning in speech in the second year. By focusing on how infants, from the beginning, attend to multiple sources of information in making sense of spoken language, this approach emphasizes the continuity between current psycholinguistic theories about how language is used and emerging developmental perspectives on how language is learned.

D. Learning about the Sounds of Speech in the First Year

To begin making sense of speech, infants must discern regularities in the sequences of sounds used by speakers of the particular language they are hearing. Hundreds of exper- iments on speech perception in the first year of life have shown that months before understanding or speaking a single word, infants become attuned to characteristic sound patterns in the ambient language (see Jusczyk, 1997). While early research in this area focused on discrimination and categorization of consonants and vowels, more recent studies are exploring infants’ implicit learning of complex distributional patterns in spo- ken language and how these learning strategies enable infants to find the words, using multiple sources of information available in the sound patterns of continuous speech.

1. Early Attention to Speech Sounds

Given the abundance of findings in what is now regarded as one of the most exciting areas of research in cognitive development, it is hard to believe that infants’ sophisticated speech-processing abilities are a relatively recent discovery. To give some historical per- spective, in 1970 Bernard Friedlander published a research overview entitled Receptive Language Development in Infancy: Issues and Problems, motivated by the following concerns: Judging by the theoretical and speculative literature as it stands today, receptive language development in infancy is a minor topic of marginal significance. Issues related to infant listening and receptive processes are virtually ignored in ... the new wave of language studies that assumed torrential proportions in the early 1960's. Though there is a general acknowledgment.... that language input is a necessary prerequisite for the organization of speech, the topic is seldom accorded more than a few sentences.. . and some of these discussions are highly patronizing in tone. They seem to suggest that auditory perception in general and language perception in particular are topics on which thoughtful observers would hardly need to spend much time. There is little in this literature to suggest that the problem of how babies come to recognize the phonological, lexical, semantic, and grammatical systems in the language they hear represents a psychological, linguistic, and developmental problem of the greatest magnitude (1970, p. 7).

Friedlander's (1970) paper was more a lament than a review, because there was almost no research available at the time: "Hardly enough is known at a factual level about early listening processes and their role in language growth even to organize the phenomena in reasonably durable categories" (p. 8). Then just one year later the situation changed, when the first reasonably durable categories were discovered through the pioneering experiments of Eimas, Siqueland, Jusczyk, and Vigorito (1971). Using an innovative operant technique called the high-amplitude-sucking procedure, these researchers were able to show for the first time that young infants can discriminate and categorize speech sounds. To fluent English speakers the speech sounds /b/ and /p/ sound clearly distinct, so it is difficult to appreciate how formidable this task could be to a linguistically inexperienced listener. In fact, these two consonants are acoustically very similar; moreover, /b/ and /p/ vary acoustically when combined with different vowels or when they occur in different positions in a word. Eimas et al. showed that even very young infants could discriminate /b/ from /p/, and just like adults, they failed to distinguish different tokens of /b/ that were acoustically distinct yet were members of the same phonetic category. Infants can also appreciate the fact that vowel tokens that are acoustically dissimilar may be equivalent in terms of their phonetic identity. Using an operant head-turn procedure, Kuhl (1979) showed that five-month-old infants readily discriminated /a/ from /i/ when spoken with the same intonation by the same female speaker. However, they grouped together several different tokens of /a/ that were acoustically variable, produced by male and female speakers using both rising and falling pitch contours. These classic studies on early speech processing abilities showed that infants can attend to the acoustic variability rel- evant to the phonetic identity of speech sounds, while ignoring acoustic variability that is linguistically irrelevant.

The first experimental studies on early speech perception focused on infants' ability to distinguish isolated syllables. The questions initially of interest derived from controver- sial issues in research on adult speech perception, with the infant representing the "initial state," or the listener innocent of experience. Studies with infants were seen as test cases relevant to current debates about which acoustic features are most critical for human speech perception (e.g., Eimas & Corbit, 1973), whether speech and non-speech sounds are processed in fundamentally different ways (e.g., Jusczyk, Rosner, Cutting, Foard, & Smith, 1977), and whether speech sounds are represented in terms of phonetic features or syllables (Bertoncini, Bijelijac-Sabic, Jusczyk, Kennedy, & Mehler, 1988). While providing valuable information about early perceptual abilities crucial for speech processing, most of these early studies were "developmental" only in the sense that they showed these capabilities were already present at birth. A more dynamic picture has emerged in recent years as researchers have begun to focus on developmental change in speech processing strategies, first by exploring how experience with a particular language shapes perception, and second by asking how infants learn to recognize patterns in speech that may help them identify linguistic units.

2. Becoming a Native Listener

Although infants are clearly born with perceptual abilities and biases that equip them for organizing speech sounds into linguistically relevant categories, these perceptual grouping strategies are neither unique to humans nor unique to speech sounds. Other primates also organize human speech sounds categorically, and some other kinds of acoustic stimuli are perceived in a similar fashion (see Kuhl, 2004). What is presumably unique to humans is the perceptual learning that occurs over the first few months of life as a result of hearing a particular language. Adults often find it difficult or even impos- sible to distinguish certain speech sounds in an unfamiliar language. For example, native speakers of Hindi can easily discriminate the consonants /Ta/ and /ta/, but to monolingual English-speaking adults they sound like indistinguishable tokens from the English cate- gory /t/. However, six-month-old infants growing up in English-speaking families can effortlessly discriminate the Hindi contrast /Ta/ - /ta/ (Werker & Tees, 1984). Studies of adult perception of native and non-native speech sounds show that adults have become specialists, attentive to phonetic distinctions relevant in the languages they have learned but less discerning in making other distinctions. Yet infants must start out with the poten- tial to make a wide range of distinctions. When does this process of perceptual specialization begin? Werker and Tees tested English-learning infants at three ages between 6 and 12 months, to investigate whether they retained their ability to discriminate non- native speech contrasts across the first year.

Infants at each age listened either to the Hindi consonants /Ta/-/ta/ or to consonants from the Nthlakampz language, /k'i/-/q'i/, which are also very difficult for English-speaking adults to discriminate. Almost all of the infants at 6–8 months could discriminate both non-English contrasts, although very few of the infants at 10–12 months were able to distinguish either pair. Further evidence for the influence of the ambient language on infants' emerging pho-netic categories comes from research by Kuhl, Williams, Lacerda, Sterens, & Lindblom (1992), who showed that six-month–old infants hearing only Swedish or English already grouped vowels perceptually in categories appropriate to the language they were learn- ing. Recent studies measuring brain activity are generally consistent with the behavioral findings showing increasing specialization for familiar speech sounds over the first year. At six months of age, infants show an electrophysiological response to changes in both native and non-native speech contrasts, but by 12 months the response is elicited only by changes in speech sounds native to the language the child has been hearing (e.g., Cheour- Luhtanen et al., 1995). These results indicate that auditory experience over the first-year results in neural commitment to a particular perceptual organization of speech sounds appropriate to the ambient language. Through early experience with the speech around them, infants adapt their perceptual strategies for efficiency in processing the language they are learning.

3. Finding the Words in Fluent Speech

Other studies of developmental change in speech perception have focused on the dis- covery procedures infants use to identify higher-order elements in spoken language. An influential article by Lila Gleitman and colleagues stimulated this new research direction (Gleitman et al., 1988). They proposed that infants might be able to use certain prosodic features in continuous speech, such as pauses and the vowel lengthening typically preceding pauses, as cues to the boundaries of phrases and clauses, a perceptual discovery strategy that could be useful to the child beginning to learn syntax. This "prosodic boot- strapping hypothesis" generated considerable interest, leading to experiments showing that 10-month-old infants seemed to recognize violations of common prosodic rhythms in the ambient language (e.g., Kemler Nelson, Hirsh-Pasek, Jusczyk, & Cassidy, 1989). Although there were also counterarguments against the view that prosodic cues are sufficiently regular as to provide reliable cues to syntactic units in speech (e.g., Fernald & McRoberts, 1995), the prosodic bootstrapping hypothesis stimulated the first wave of research exploring how infants might use lower-level acoustic cues in speech to gain access to linguistic structure at higher levels (see Morgan & Demuth, 1996). The strong claim that young infants must first rely on prosodic information in order to organize segmental information in continuous speech has receded in light of new find- ings showing that infants are much more adept at identifying word-like units in fluent speech than anyone had imagined. Jusczyk and Aslin (1995) investigated the ability of seven-month-old infants to detect repeated words embedded in fluent speech. When in- fants were first familiarized with multiple repetitions of a word such as bike or feet and then tested in an auditory preference procedure with passages that either did or did not contain the familiarized word, they preferred to listen to passages containing the familiar word. This finding indicated that infants were able to segment speech into words without benefit of exaggerated prosodic cues. However, prosody at the level of word stress does play a role in facilitating such segmentation. English-learning infants are more success- ful in segmenting words such as bor'der that have a strong–weak accent pattern than words such as guitar' that have the opposite pattern, because they have already learned that the strong–weak pattern is dominant in the language they are hearing (Jusczyk, 1998). In contrast, French-learning infants appropriately show the opposite bias, based on their experience hearing words with weak–strong accent patterns.

Many studies have now demonstrated infants' sensitivity to particular cues in the ambient language such as phonotactic regularities (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993) and lexical stress (Jusczyk, Cutler, & Redanz, 1993), and their ability to take advantage of these cues in identifying word boundaries (e.g. Johnson & Jusczyk, 2001). In an influential study mentioned earlier, Saffran et al. (1996) showed that infants were also able to use sequential statistics to discover word-like units in con- tinuous speech, in the absence of any other acoustic cues to word boundaries. Unlike the experiments on segmentation by Jusczyk and Aslin (1995) in which infants were first familiarized with samples of natural language, Saffran et al. exposed infants briefly to strings of meaningless syllables, i.e., stimuli that were language-like but entirely novel in their organization. After only a few minutes of passive exposure to these sounds, infants picked up on the regularities and attended longer during testing to sequences that devi- ated from these regularities. Using a similar training procedure, Chambers, Onishi, and Fisher (2003) showed that infants can also learn new phonotactic regularities after mini- mal exposure. Moreover, they can also quickly take advantage of such newly learned phonotactic patterns, using them as cues to identify the boundaries of novel words (Saffran & Thiessen, 2003). Thus although infants in the second half of the first year may already show a strong commitment to the particular sound patterns they have absorbed from hearing their native language, early speech processing remains a highly dynamic process. Infants remain open to new experience as they build on prior learning, drawing on multiple sources of information to find order in novel sounds.

Experimental studies of early speech processing proceed parametrically, typically investigating one isolated variable at a time. However, infants listening to natural speech are in fact confronted with multiple sources of information at any moment, some redundant and others in conflict. Building on a strong foundation of research on infants' use of individual cues to word boundaries, several recent studies have begun to explore how young language learners tackle this challenging problem. Just as research on adult speech processing now focuses on how listeners integrate probabilistic information from numerous sources (Seidenberg, 1997), developmental researchers are beginning to ask parallel questions of very young infants (e.g., Curtin, Mintz, & Christiansen, 2005; Mattys, White, & Melhorn, 2005; Thiessen & Saffran, 2003, 2004). These and many other new findings confirm the wisdom of Friedlander's (1970) intuition 35 years ago, "that the problem of how babies come to recognize the phonological, lexical, semantic, and grammatical systems in the language they hear represents a psychological, linguistic, and developmental problem of the greatest magnitude" (p. 7).

E. Lexical Development

These studies of early speech perception show that infants in the first year of life are becoming skilled listeners, capable of making detailed distributional analyses of acoustic–phonetic features of spoken language. Although such accomplishments are often cited as evidence for early "word recognition," they are perhaps more appropriately viewed as evidence of pattern detection abilities pre-requisite for recognizing words in continuous speech. Identifying particular sound sequences as coherent acoustic patterns is obviously an essential step in word recognition, but this can occur without any asso- ciation between sound and meaning. Laboratory studies show that around 5–6 months of age infants respond selectively to their own name (Mandel, Jusczyk, & Pisoni, 1995), and by 10 months appear to have some kind of acoustic–phonetic representation for a num- ber of frequently heard sound patterns (e.g., Halle & de Boysson-Bardies, 1994). Because this selective response to familiar words in the early months of life can occur with no evidence of comprehension, it may constitute word recognition in only a limited sense. However, most infants do begin to respond to and utter sounds in meaningful ways by their first birthday, and one year later are able to speak dozens of words quite con- vincingly. In this section, we review research on children's first speech productions and the course of early vocabulary growth, as well as the factors that influence word learning in infancy.

1. First Words

According to parents' reports of their children's spontaneous responses to speech, infants typically begin to associate sound sequences with meanings toward the end of the first year. By eight months, on average, many children respond appropriately to about 10 familiar phrases, such as "Where's Daddy?" (e.g., by turning and crawling toward the door), or "It's time for bath!" (e.g., by plopping down and attempting to remove their shoes) (Fenson et al., 1994). While it could be tempting to assume that the child is actu- ally interpreting each of the words in these phrases, it is more likely that children are using a variety of cues, both linguistic and contextual, to make sense of these frequent expressions (e.g., Daddy just left the room, Mom is holding a towel standing next to a running faucet). At the same time, the fact that children do respond in these ways indi- cates that they are paying attention to the speech around them and beginning to under- stand it by associating certain sound patterns with particular contexts. Only a short time later, children begin to demonstrate an ability to understand individual words with less and less contextual support, an ability that will continue to improve over the next several months. Based on reports from more than 1000 parents, Fenson et al. (1994) cite that the median-level 10-month-old understands approximately 40 words, while the median-level 18-month-old understands more than 250 words, a more than six-fold increase. The production of recognizable words begins, on average, just before a child's first birthday. These early words cross-cut a variety of linguistic categories, but are typically names for caregivers (e.g., mama), common objects (e.g., bottle, shoe), social expres- sions (e.g., bye-bye), with some modifiers (e.g., hot) and actions or routines (e.g., peek- aboo, throw) (Nelson, 1973). New words tend to enter children's expressive vocabularies over the next several months at a relatively slow but steady pace, reaching an average of 300 words by 24 months and more than 60,000 by the time they graduate from high school (Fenson et al., 1993). Thus, after the slow start, many children appear to undergo a "vocabulary burst," a sudden and marked increase in how many words children use (e.g., Goldfield & Reznick, 1990; Mervis & Bertrand, 1995). Putting aside the difficul- ties of defining how much of an increase constitutes a "spurt" in rate of learning, the char- acterization of lexical development in terms of a sudden increase in learning rate has been interpreted to indicate two distinct phases of lexical acquisition (i.e., pre-spurt vs. post- spurt). Some researchers have associated this "burst" with the achievement of linguistic milestones, for example, children's new understanding about what words are for (the "naming insight") (e.g., Dromi, 1987; Bloom, 1973), improved word segmentation abi- lities (Plunkett, 1992), or enhanced word retrieval skills (Dapretto & Bjork, 2000). Other researchers have associated increases in lexical growth with cognitive advances related to the nature or organization of object concepts (e.g., Gopnik & Meltzoff, 1987). However, other researchers have suggested that these increases in the rate of learning are relatively constant across the period, questioning the idea of a "spurt" at all, and hence, its role as a marker for other cognitive or linguistic events (Bates & Goodman, 1999; Bloom, 2000; Ganger & Brent, 2004). Moreover, work within a connectionist framework suggests that it may not be necessary to assume that any shift in the trajectory of vocabulary growth would be associated with the emergence of a new insight or learn- ing mechanism at all. Instead, both the slow and steady pace of learning early on and the accelerated learning later can be accounted for within a single explanatory framework. In Plunkett et al. (1992) a connectionist network was trained to associate labels (i.e., words) to random-dot "images" (i.e., pictures). In the "language" and "world" of this network, several images had the same "name" and there was no information in the images regard- ing what the label should be. Just like in natural languages then, the mappings between labels and images were arbitrary and many-to-one. Importantly, however, in the network, a vocabulary "spurt" occurred without any shift in the underlying mechanism guiding the learning. That is, even though the identical network was solving an identical task throughout development, shifts in learning rate were observed in the behavior of the net- work. As summarized in Elman et al. (1996), "there is no need to build in additional architectural constraints or to invoke changes in the input to explain the vocabulary spurt. It is simply an emergent function of the processing" (p. 128). The shifts were a natural consequence of processing limitations that arise over the course of a gradual and conti- nuous learning process that involves complex and multiply determined mappings.

The fact that learning words involves gradually building a system of mappings using a variety of linguistic and non-linguistic constraints is also evident when examining changes in the types of words that children produce. Even though children's first words come from a range of lexical categories (e.g., Bloom, 1973), first words are typically open class or content words (nouns, verbs, adjectives), and only later grammatical func- tion words such as prepositions, determiners, and pronouns. Within the open class, chil- dren's first words tend to be referential (i.e., concrete nouns), and only later are children producing predicative terms (e.g., verbs and adjectives) (Bates, Bretherton, & Snyder, 1988; Benedict, 1979; Brown, 1973; Nelson, 1973). This dominance of concrete nouns in early vocabularies (sometimes referred to as a "noun bias") is most evident in the first 200 words or so, after which there tends to be an increase in the proportion of vocabu- laries devoted to predicative terms, for example, verbs (e.g., go), adjectives (e.g., hot), and closed-class functors (e.g., and, of) (Dromi, 1987; Nelson, 1973; Bates et al., 1994). Several explanations have been proposed for these developments. First, it is possible that open-class words are learned early because they are longer in duration, generally stressed, and phonologically less reduced than closed-class words (Morgan, Shi, & Allopenna, 1996). Even six-month-old infants show a preference for listening to open class rather than closed-class words (Shi & Werker, 2001), and newborns are sensitive to the acoustic differences between these word types (Shi, Werker, & Morgan, 1999). Second, it might also be the case that the ability to learn predicates is dependent on amassing a particular body of referential terms, and that using words that do grammatical work (e.g., functors) is dependent on the acquisition of a set of content words on which they can operate (i.e., "from reference to predication to grammar," Bates et al., 1994, p. 98). Third, concrete nouns are conceptually simpler than verbs and other predi- cates, and both of these types of open-class words are more conceptually transparent than grammatical function words (Gentner, 1982; Gasser & Smith, 1998).

More specifically, it has been proposed that the early priority of nouns in children's vocabularies reflects the fact that nouns are more "cognitively dominant" than verbs and closed-class items (Gentner, 1982; Gentner & Boroditsky, 2001). That is, nominals serve primarily to denote concrete objects that are more bounded and more perceptually individuated than, for example, verbal terms denoting states and processes which must rely on their argu- ments (i.e., nouns) to make sense. Relational words (e.g., closed class terms), in contrast, are "linguistically dominant," in that they derive their meaning from other parts of the linguistic context and sometimes reflect relatively opaque grammatical constructs (e.g., gender) (Gentner & Boroditsky, 2001). The general pattern of early nominals appearing before other more relational terms has been observed in the vocabularies of children from several language-learning communi- ties, even though their languages have different typological features that potentially make them more or less "noun friendly" (Gentner, 1982; Caselli, Casadio & Bates, 1999; Jackson-Maldonado et al., 2003; Bornstein, et al., 2004). For example, Caselli, Casadio, and Bates, (1999) used parent report to contrast early vocabulary composition in young English- and Italian-learners. In spite of the fact that Italian, but not English, is a pro-drop language (i.e., the subject noun phrase can be omitted leaving verbs in the very salient sentence-initial position), a similar level of noun bias was observed in the two languages. Likewise, in a recent study of middle-class children learning English, Italian, Spanish, Dutch, French, Hebrew, and Korean, Bornstein et al. (2004) report few crosslinguistic differences in vocabulary composition, with mothers of children in all language commu- nities reporting more nouns than words in other language classes. Interestingly, Caselli et al. also report that Italian-learning children were likely to have more social terms and names for people in their vocabularies than English-learners, suggesting that cross- linguistic differences in vocabulary composition may be more attributable to social or cultural factors (e.g., a tendency to live near extended family) than specific features of the language. Similarly, in a study of children learning English, Italian, and Spanish in urban and rural communities (Bornstein & Cote, 2005), the observed variation in vocabulary size and composition was generally attributable to cultural factors, favoring urban over rural settings, rather than to the particular language being learned.

However, the claim that a "noun bias" may be a universal feature of early conceptual and linguistic development has been challenged by other studies of children learning Korean (Choi & Gopnik, 1995), Mandarin Chinese (Tardif, Gelman, & Xu, 1999), and Japanese (Fernald & Morikawa, 1993). All of these languages have features that make them more verb-friendly than English and, Italian. For example, while both Italian and Mandarin are "pro-drop" languages, Mandarin verbs are considered to be more morpho- logically transparent than Italian verbs with their rich morphology. Indeed, naturalistic observational data indicated that young Mandarin-learners produced a higher proportion of verbs than nouns, compared to their Italian- and English-speaking peers (Tardif et al., 1999). This is in contrast to a consistent pattern of noun dominance using parent report, suggesting that parent report may tend to over estimate nouns (and underestimate verbs) in children's vocabularies.

Interestingly, Tardif et al. (1999) found that Mandarin moth- ers produced more verbs than nouns in their spontaneous speech to their children, and Fernald and Morikawa (1993) found that Japanese mothers labeled objects less fre- quently and less consistently than English-speaking mothers. In several studies, English- speaking mothers were more likely to use and elicit more nouns than verbs from their children, and place them in salient positions in the sentence when engaged in activities such as object-naming or book-reading (e.g., Hoff, 2003; Tardif, Shatz, & Naigles, 1997). Again, it appears unlikely that a single factor can account for the "dominance" of nouns in children's early vocabularies, but rather children's early vocabulary compositions are determined by a multitude of factors that happen to vary across languages and language- learning situations.

While many studies have examined the early stages of vocabulary development, it is still difficult to ascertain what a child really knows when they understand or produce a word. When talking about early lexical development, it is tempting to credit a child with "knowing" a word, as if word knowledge is something that the child either has or does not have, i.e., as if words are acquired in an all-or-none fashion. However, early studies noted that children's early words (e.g., bottle) do not necessarily have the same meanings (e.g., white plastic 6 oz cup with the bright red screw-on lid) as they would for the adult (e.g., receptacles of all shapes and sizes from which one generally pours liquids). Early words are frequently used in very context-specific ways (i.e., under-extensions), only with reference to specific objects in specific situations (Bloom, 1973; Barrett, 1986; Harris, Barrett, Jones, & Brookes, 1988). At the same time, children's early word uses might also be considerably more broad (i.e., overextensions) than one would expect based on adult-like meaning categories (Bates, Benigni, Bretherton, Camaioni, & Volterra, 1979). For example, a child using the word "dog" to refer to all four-legged animals (dogs, but also cats, cows, and horses) could be an example of a child misrepre- senting the adult-like meaning for words. Yet, research has also shown that over- or under-extensions are actually relatively rare (Rescorla, 1980; Harris et al., 1988; Clark, 2003b), and may be less of a reflection of how children represent the word's meaning in some sort of mental lexicon and more related to the child's ability to put their lexical knowledge to work in real time (e.g., Huttenlocher & Smiley, 1987).

That is, a child who uses the word "dog" when the family cat runs across his path may not actually think that dogs and cats are the same thing, but simply cannot generate the appropriate word in the heat of the moment. Interestingly, children's over- and under-extensions are considerably more frequent in production than comprehension (e.g., Clark, 2003a), and an experimental study using a looking-preference procedure has shown little concordance between comprehension and children's over- and under-extensions in production (Naigles & Gelman, 1995). Taken together with research on children's processing of speech in real time, described in more detail below (e.g., Fernald, Perfors, & Marchman, 2006), these studies suggest that early lexical development is quite gradual. It involves not only build- ing-up "adult-like" meaning representations, but also learning to use words in more and more contextually flexible ways and in more and more challenging contexts (Bates et al., 1979; Barrett, 1976).

2. Individual Differences in Vocabulary Development

So far, we have been talking about the general features of lexical development in "the modal child" (Fenson et al., 1994, p. 1). However, there is considerable variation in both when and how children build their receptive and expressive vocabularies (Bates et al., 1988, 1994; Bloom, Lightbown & Hood, 1975; Fenson et al., 1994; Goldfield & Snow, 1985; Nelson, 1973, Peters, 1977, 1983). For example, while many children show signs of word comprehension at 8 or 10 months of age, other children do not respond systemati- cally to the speech around them until several months later. Similarly, some children pro- duce their first words well before their first birthday, while others do not do so until 14 or 15 months of age. The "modal" 18-month-old has already built up a 50–75 word expres- sive vocabulary, yet other children do not amass this many recognizable words until 22 months or later. Some of these "late talkers" will catch up in vocabulary a few months down the road, while others will remain late and continue to be at risk for language or learning disorders (Bates, Dale, & Thal, 1999). Studies of variation in vocabulary compo- sition have noted that some English-speaking children tend to adhere to a strong "noun bias" tendency, a so-called "referential style" of early word learning (e.g., Nelson, 1973; Pine & Lieven, 1990). In contrast, other children with a more "expressive" style tend to have a smaller proportion of concrete nouns, preferring more "canned phrases" (e.g., I wanna do that!) and social expressions (e.g., “no way Jose!”) (Nelson, 1973). Such individual differences have been well-documented since the mid-1970s, based pri- marily on diary studies (e.g., Nelson, 1973). However, important progress in understanding the extent of individual variation has been facilitated by large-scale studies which rely on reports from parents, e.g., MacArthur-Bates Communicative Development Inventories (CDI) (Fenson et al., 1993) and the Language Development Survey (Rescorla, 1989).

While there are clearly limitations to this methodology (e.g., Mervis & Tomasello, 1994), this technique has enabled the examination of variation in lexical milestones in several lan- guages, for example, English (Bates et al., 1994), Italian (Caselli et al., 1999), Mexican Spanish (Jackson-Maldonado et al., 2003), Hebrew (Maitel, Dromi, Sagi & Bornstein, 2000), as well as in children learning two languages simultaneously (Pearson, Fernández & Oller, 1993; Marchman & Martínez-Sussmann, 2002) and in children from urban and rural settings (Bornstein & Cote, 2005). Although there is clearly variation in early acquisition across languages (e.g., Caselli et al., 1999; Choi & Bowerman, 2001; Tardif et al., 1999), research has consistently demonstrated remarkable similarities across languages in the overall size of children’s vocabularies and the extent of the variation that is observed. Commenting on their comparative data across Spanish, English, and Italian, Bornstein and Cote (2005) noted a strikingly similar range of vocabulary knowledge in all three lan- guages, leading these authors to conclude that individual variability is probably a universal feature of early language acquisition. What are the sources of individual differences so early in development? Some studies have looked to child-factors, such as gender or birth order, to explain variation in lexical development. Studies have documented somewhat larger vocabularies and faster rates of growth in girls compared to boys (Fenson et al., 1994; Huttenlocher, Haight, Bryk, & Seltzer, 1991) and first-borns compared to later-borns (e.g., Hoff-Ginsberg, 1998). Recently, Bornstein and Cote (2005) note a consistent advantage for girls over boys in re- ported vocabulary in English, Italian, Spanish, across both urban and rural communities. While it is striking to see consistent gender effects in communities that likely vary in gen-der-based social expectations, these effects are generally small relative to overall devel- opmental effects (i.e., the impact of gender is considerably smaller in magnitude than age effects). Like gender, the impact of birth order is relatively minor compared to other fac- tors, but points to the suggestion that children, even those living in the same family, can differ in the frequency and character of interactions in which they engage on a regular basis. Indeed, it is well-known that there are considerable individual differences in the quantity and quality of the talk that children hear (e.g., Huttenlocher et al., 1991; Huttenlocher, Vasilyeva, Cymerman, & Levine, 2002; Hart & Risley, 1995), and that several of these features of maternal talk are directly linked to children's vocabulary out- comes (Hoff, 2003).

In a recent large-scale study of low-income families, Pan, Rowe, Singer, and Snow (2005) found that variation in growth in children's vocabulary from 14 to 36 months was significantly related to diversity of maternal talk, in particular, the num- ber of different words produced during mother–child interaction. Thus, children who hear a rich vocabulary that includes a higher proportion of low-frequency or complex words are likely to develop their own vocabularies at a faster rate (see also Weizman & Snow, 2001; Hoff & Naigles, 2002). However, Pan et al. (2005) also found that features of maternal knowledge (e.g., scores on standardized tests of language and literacy) and maternal mental state (e.g., depression index) also contributed to child outcomes. Other researchers have characterized individual differences in terms of cognitive or processing abilities. Typically based on naturalistic data, several early studies proposed that children may vary in the tendency to select analyzed vs. unanalyzed units (e.g., Peters, 1983), the use of strategies for segmentation which favor "word-sized" vs. "phrase-sized" units (Plunkett, 1992; Bates et al., 1988), a predilection for "imitative- ness" (Bloom et al., 1975), or the ability to use contextual or linguistic cues to retrieve words (Bloom, 1973). Other studies have used more processing-based measures to assess various skills that could underlie vocabulary development. Using a phonetic discrimina- tion task, Werker, Fennell, Corcoran, and Stager (2002) found that 14-month-olds who were successful at learning phonetically similar words had relatively larger vocabularies. Interestingly, no relation was observed in 18-month-olds, suggesting that this type of skill may only be helpful at the beginning phases of building a vocabulary. Similarly, Swingley, and Aslin (2000, 2002) also found little relation between vocabulary size and the ability of 14- and 18-month-olds to identify words pronounced correctly and incor- rectly. However, Fernald and colleagues (Fernald et al., 2001; Fernald, 2002) found that 18- and 21-month-olds who had larger production vocabularies were faster than their lower-vocabulary peers to recognize words based on partial phonetic information and were more efficient at using verb semantics to predict what was coming up in the sen- tence.

Using a similar procedure, Zangl, Klarman, Thal, Fernald, and Bates (2005) found that infants with larger vocabularies were more efficient at processing words that were perceptually degraded. Finally, in a longitudinal sample, Fernald et al. (2006) have re- cently shown that efficiency of spoken language understanding was related to trajectories of growth in vocabulary from 12 to 25 months, as well as several indices of early gram- mar. Thus, individual differences in children's burgeoning vocabulary knowledge appear to be linked to a variety of skills that come into play during the processing of both linguis- tic and non-linguistic information during real-time language comprehension. While we are still a long way from knowing exactly how those factors operate over the course of development, it is likely that individual differences in lexical development are linked to a host of factors, both child-related and experience-related, that all contribute to the vari- ation that is so pervasive in vocabulary development.

3. Early Word Learning

Observational studies and research using parental report measures can provide norma- tive data on the rate and composition of vocabulary growth by children at different ages, as well as correlational data showing how environmental factors such as the amount and quality of parental speech relate to lexical development (e.g., Huttenlocher et al., 1991). However, an experimental approach is required to examine how young children make use of particular sources of information in the process of figuring out the meanings of novel words. For example, imagine a scene where a two-year-old visiting her relatives is served an unfamiliar fruit pastry after dinner, and her mother exclaims "What a surprise! Rhubarb!" How is the child to make sense of this remark? Studies exploring versions of this question now number in the hundreds (see Bloom, 2000; Woodward & Markman, 1997). Here we provide a very brief overview of research on some of the factors influ- encing young children's interpretation of novel words. Discussions of this research question often start by framing the problem in terms of Quine's (1960) vignette of a linguist visiting an unfamiliar culture, accompanied by a native guide who does not speak his language. As a rabbit hops across the scene, the guide exclaims Gavagai! How, Quine asks, can the stranger possibly interpret this utter- ance, given that the speaker could be referring to the rabbit, to a part of the rabbit, to the animal's action, or to an indeterminate number of other aspects of the scene? Quine does not provide an answer to this conundrum, emphasizing instead the fundamental impossi- bility of knowing the intended meaning based on the evidence at hand.

In the language development literature, however, it is assumed that a young child in this situation would most likely rapidly and automatically interpret gavagai as rabbit, without considering the myriad other possibilities. Thus the inherent indeterminacy of meaning that was the focus of Quine's argument is circumvented by the young language learner thanks to interpretative biases that guide early word learning. The idea that word learning gets started with help from some sort of "object-category bias" is supported by experiments in which chil- dren are asked to consider novel words in relation to novel objects. When an unfamiliar object such as a toy animal is labeled with an unfamiliar name (e.g., ferret), children typically assume that the new word refers to the animal as a whole, rather than its tail, its color, or the stuff it is made of (e.g., Markman & Hutchinson, 1984).

Although a bias for naming the whole object is predictably observed in experiments of this sort, there is considerable debate as to how this phenomenon should be interpreted. What factors lead the child to guess that the new word ferret is a name for the animal as a whole and other animals of the same kind, rather than its parts or properties? Three dif- ferent approaches to this question have been discussed extensively: The first approach emphasizes the importance of preverbal perception and cognition in guiding word learn- ing. Because objects are perceived as bounded and coherent and thus are salient even to infants (e.g., Spelke, 1998), we are predisposed from infancy on to see the world as con- taining cohesive objects. This could explain why children as well as adults are biased to identify (and to name) an object as a whole before attending to its parts and other attrib- utes (Gentner, 1982). Moreover, nouns naming concrete objects are conceptually simpler than relational words like verbs and adjectives, which can vary substantially in the per- ceptual features they refer to depending on the nouns they are associated with (e.g., a good cookie vs. a good dog). According to Gentner and Boroditsky (2001), such non- linguistic aspects of human perception and cognition in relation to language structure could account for the tendency of infants to learn names for objects before they learn names for actions and attributes. A second approach to studying factors that guide early word learning emphasizes the critical role of social cognition, a perspective that has roots in Vygotsky's (1962) theory of social support for learning and in Bruner's (1975) views on how reference first emerges in preverbal communication. Current research on the social origins of linguistic knowledge focuses on children's emerging awareness of the referential intentions of others in figuring out what unfamiliar words might refer to.

For example, several studies have shown that when children hear a novel word, they will connect it with an appropriate object only if they somehow appreciate that the adult intended to name the object (e.g., Baldwin et al., 1996; Tomasello & Barton, 1994). A third approach to understanding early word learning proposes linguistic constraints that account for children's biases in interpreting novel words. According to Markman's (1989) influential formulation of this position, such constraints are default assumptions that serve to limit hypotheses as to possible meanings for a new word. In particular, the whole object assumption guides the child to associate a new word with the entire object, while the taxonomic assumption guides the child to extend that new word to other objects belonging to the same class rather than to thematically related objects. Another proposed word learning constraint is the mutual exclusivity assumption, a kind of exclusionary learning strategy that has recently been demonstrated in studies with infants as young as 15 months of age (Halberda, 2003; Markman, Wasow, & Hansen, 2003). When a young child is presented with a familiar object with a known name (e.g., a ball) along with an unfamiliar object (e.g., a whisk) and is then asked to find the dax, the typical response is to choose the whisk. One interpretation of this effect is that the child automatically maps the novel word onto the novel object rather than assigning a second name to the ball, guided by the default assumption that an object can only have a single name (e.g., Markman & Wachtel, 1988). Other researchers disagree that this effect is specific to word learning, arguing instead that it is grounded in pragmatic knowledge (see Bloom, 2000; Clark, 2003b). For example, Clark (1997) argues that children's bias against lexical over- lap is best explained in terms of a "principle of contrast" which leads them to assume that differences in form should correspond to differences in meaning. It is interesting to note that even dogs show a related form of exclusionary learning, mapping novel spoken words onto objects for which no name has previously been learned (Kaminski, Call, & Fischer, 2004). However, although this finding suggests that a learning principle based on mutual exclusivity might not be specific to human language, it is also clear that animal learning of word–object associations differs in important ways from lexical learning by children (Bloom, 2004; Markman & Abelev, 2004) The robust learning biases demonstrated in these experiments are certainly consistent with the notion that children rely on strategies specifically adapted for lexical learning. However, the view that such learning biases are automatic and language-specific has come under criticism from many directions. Children's early vocabularies do not consist only of nouns and include types of words (Hi! Up! More!) quite different from the object names supposedly favored by lexical constraints (Nelson, 1973).

For some first words (e.g., bath) that are technically nouns, it is not at all clear whether they are under- stood by the child as an object or an action, or as a routine involving both. Although word learning constraints are proposed as a solution to the problem characterized in Quine's (1960) dilemma, many scholars of early language learning (e.g., Bloom, 1993, 2000; Clark, 2003a, 2003b; Nelson, 1988; Tomasello, 2003) point out that the young child is not really comparable to the linguist who speaks a first language different from that of the native guide, and that the guide is not at all like the parent of a small child. In the ecology of early parent–child interaction, the adult takes the perspective of the linguistic novice in many ways, providing both simplified language input (e.g., Lieven, 1994) and many pragmatic cues to reference using gaze and gesture that young children are able to use in figuring out what new words mean (Tomasello, 2003). Another way in which Quine's example of linguistic indeterminacy has been oversim- plified as a model of early word learning is that novel objects do not always hop across the scene as the single most salient focus of attention. Imagine a variation of this sce- nario: the native guide fires his gun as he yells Gavagai! and the rabbit falls down dead. Under these circumstances, the linguist as well as the child would presumably not default automatically to a static whole-object interpretation of this string of sounds, but might as- sume Gavagai! meant something like Watch out! or Got it! or Dinner! The everyday sit- uations in which word learning occurs are often much more dynamic and ambiguous than the experimental setups in which word learning is studied scientifically.

In the example above in which the mother exclaims "Rhubarb!" as the dessert is placed on the table in front of her 2-year-old, the child will only gradually figure out what this word refers to, learning based first on the taste and texture of the cooked fruit and only later on other properties of rhubarb as a plant. Does this mean that a mapping error will occur, as the child automatically attempts to apply the new word to the pastry itself as a whole object? It seems more likely that the child will pick up on social cues indicating that the mother's remark is addressed to others at the table and is not intended as a label, one of several reasons a word may simply be ignored on first exposure. After all, infants hear thousands of words in a week yet learn to use only one or two new words a day. Although critics of the linguistic constraints position may object that the point of Quine's (1960) example has been distorted in the developmental literature, they agree that young children face a daunting inductive problem in assigning meaning to an unfa- miliar sequence of speech sounds. But it is possible to agree that children need to limit the potentially large number of hypotheses for word meanings without assuming either that this kind of inductive problem is unique to word learning, or that constraints in the form of default assumptions are the only way to solve the problem.

Bloom (2000) points out that children face comparably complex inductive problems in other domains of expe- rience all the time. When a child grabs the handle of an iron skillet sitting on a hot stove burner, she has to figure out what to avoid in the future in order not to get burned again. Is it the skillet or just the handle? Or could it be anything shiny and white like the stove? In this case, avoidance of the skillet itself might reflect a non-linguistic whole-object bias, a reasonable first guess until the child developed a deeper understanding of the causal processes involved. In another example of a proposed word-learning constraint that may be much more general in its scope, Markson and Bloom (1997) show that the phenomenon of "fast mapping" a novel word to a novel object is not limited to lexical learning. When children hear a novel word described as a koba, they remember which object the new word referred to; however, when they hear a novel object described as "the one my uncle gave me," they are equally good at remembering which object the factnreferred to. Bloom and Markson (2001) argue that many of the findings about how novel words are extended are best explained in terms of general cognitive systems such as those involved in concept formation and intentional inference, rather than through proposed linguistic constraints that are dedicated uniquely to word learning. In an early critique of the theory that linguistic constraints are essential for learning new words, Nelson (1988) pointed out that the first formulations of this position traced their intellectual roots to Chomsky's (1975) claims for innate mechanisms for learning grammar, extending this framework to lexical learning.

This perspective is consistent with an emphasis on default learning strategies that privilege some sorts of information and are impervious to others. Just as some nativist theories of adult language compre- hension posit autonomous parsing strategies favoring syntactic structure over all other kinds of linguistic information (e.g., Frazier & Rayner, 1987), the initial emphasis in con- straints theory was on how children are restricted by strong, possibly innate, default assumptions as to what a new word might mean (e.g., Markman & Hutchinson, 1984). In more recent accounts of the linguistic constraints position, these word learning strategies are framed as more flexible heuristics, i.e., as somewhat "softer" constraints (Woodward & Markman, 1997). But the focus is still on how children are inherently limited in their interpretative strategies, rather than on how they may integrate different sources of information in different contexts.

As mentioned earlier, the idea of inflexible parsing strategies in adult comprehension has been challenged by many new studies showing how listeners integrate probabilistic information from multiple sources (Seidenberg, 1997). In research on early word learn- ing as well, there is mounting evidence that infants use diverse sources of linguistic and non-linguistic information in making sense of new words, guided by learning biases that are construed more appropriately as preferences than as constraints (e.g., Bloom, 1993; Hollich, Hirsh-Pasek, & Golinkoff, 2000). As Nelson (1988) put it, "The connotation is quite different: Constraints imply restriction – a closing down of choice; whereas prefer- ence implies free, but biased, choice" (p. 228). These new models of early word learning draw on insights articulated years ago in the "competition model" of Bates and MacWhinney (1979, 1987, 1989), namely that multiple sources of information are avai- lable to the young language learner and that the influence of different information sources varies both as a function of their strength as cues in relation to other cues, and also as a function of the developmental level of the child. Research on word learning is just begin- ning to investigate the relative contributions of multiple cues on novel word interpretation by children at different ages. For example, many studies have shown how the shape of a novel object influences children's categorization and naming.

Although in their everyday experience, children often experience new objects in motion surrounded by other objects, almost all experiments on the "shape bias" have used isolated static objects as experi- mental stimuli. However, when Smith (2005) presented 2-year-olds with dynamic stim- uli, she found that movement influenced children's judgments as to which objects were similar. Findings like this will lead developmental researchers increasingly toward a dif- ferent formulation of the question, one more in line with the probabilistic constraints approach to investigating language comprehension by adults (Seidenberg & MacDonald,1999). To get beyond the question "Is there or is there not a shape bias?" (e.g., Cimpian & Markman, 2005), studies will begin to investigate when and under what circumstances shape is an important factor in object categorization, and how shape interacts with other perceptual features as well as linguistic and social cues in the referential context in guid- ing the child's inference as to what a new word refers to.

F. Listening for Meaning in Speech in the Second Year

Infants' early progress in developing language is often charted in terms of their increasing competence in understanding, producing, and learning individual words, an ability that is arguably shared in some respects with other species (Kaminski, Call, & Fischer, 2004; Seidenberg & Petitto, 1979). However, the ability to understand and use words flexibly in combination is a critical distinguishing feature of human language. Of course, it was Chomsky (1959) who pointed out long ago that multi-word sentences are much more than individual words strung together one-by-one. Grammatical sentences are made up of units of words that vary in size and are organized hierarchically in a large, but finite, set of complex ways that are not always obvious from the surface-level order- ing of the words. While traditional views held that the language-learning child must be innately endowed with such grammatical knowledge, recent perspectives have continued to examine ways in which such proficiency can be constructed over the course of deve- lopment. In this section, we review recent research on how infants develop impressive efficiency in understanding words in continuous speech across the second year, and on how they begin to use words in combination to express increasingly complex meanings through language.

1. The Development of Efficiency in Language Understanding

To make sense of the rapidly spoken strings of words that make up the language children hear, they must learn to process fluent speech efficiently, "listening ahead" to anticipate what is coming next in the speech stream using different sources of linguistic and non- linguistic information. Many recent studies using on-line measures of comprehension with adults have shown that skilled listeners draw on multiple sources of knowledge to process speech with remarkable speed and efficiency (e.g., Tanenhaus et al., 1995). With the refine- ment of eye-tracking techniques for use with infants, it is now possible to monitor the time course of spoken language understanding by very young language learners as well. Using a looking-while-listening procedure with English-learning infants from 15 to 24 months of age Fernald, Pinto, Swingley, Weinberg and McRoberts (1998) found dramatic gains in the speed and accuracy of word recognition over the second year. In this procedure, infants look at pictures of familiar objects while listening to speech naming one of the objects. Fifteen- month-olds responded inconsistently and shifted their gaze to the appropriate picture only after the offset of the target word, while 24-month-olds were faster and more reliable, initi- ating a shift in gaze before the target word had been completely spoken. A recent longitu- dinal study following infants from 12 to 25 months found that on-line measures of efficiency in speech processing were correlated with numerous more traditional measures of lexical and grammatical development (Fernald et al., 2006). Moreover, analyses of growth curves showed that children who were faster and more accurate in on-line comprehension at 25 months were those who showed faster and more accelerated growth in expressive vocabulary across the second year. Success at word recognition in degraded speech is also correlated with vocabulary size in the second year (Zangl et al., 2005), further evidence that speech processing efficiency is related to other dimensions of early language development. A possible benefit of the increase in processing efficiency over the second year is that it enables infants to identify words more quickly based on partial phonetic information, rather than waiting until the word is complete. However, one consequence is that the young language learner is increasingly confronted with problems of temporary ambigu- ity. When Allopenna et al. (1998) presented adults with objects that included candy and a candle and asked them to Pick up the can-, they waited to hear the next speech sound before orienting to the appropriate object.

That is, they postponed their response until the final syllable of the target word made it clear which object was the intended referent. The child who hears Where's the doll? in the presence of a doll and a dog is also faced with a temporary ambiguity, given that doll and dog overlap phonetically and thus are indis- tinguishable for the first 300 ms or so. Swingley, Pinto, and Fernald (1999) found that 24-month-olds in this situation also delayed their response by about 300 ms until disam- biguating information became available. Even when they heard only the initial phonemes in familiar words (e.g., the isolated first syllable of baby or kitty), 18-month-olds were able to use this limited information to identify the appropriate referent (Fernald, Swingley, & Pinto, 2001). Further evidence for early use of phonetic information in a probabilistic fashion comes from studies by Swingley and Aslin (2000, 2002) showing that even younger infants can identify familiar words when they are mispronounced, but respond more strongly to the correct than to the incorrect version (e.g., baby vs. vaby). Children also become increasingly attentive to prosodic and morphosyntactic regularities in speech that enable them to anticipate upcoming content words in the sen- tence, also relying on probabilistic information (Fernald & Hurtado, 2006). For example, 2-year-olds expect an object name to follow an unstressed article (Zangl & Fernald, under review). When an uninformative adjective occurs instead (e.g., Where's the pretty CAR?), they "listen through" the prenominal word and wait for the noun before responding; however, if the adjective is novel and accented, they are more likely to mis- interpret the unknown word as a potential object name (Thorpe & Fernald, 2006). That is, when the word preceding the target name is stressed as well as lexically ambiguous (e.g., Where's the ZAV car?) it becomes relatively more noun-like, and 26-month-olds are more likely to respond accordingly by searching for a novel referent as soon as they hear zav, rather than waiting for the subsequent word that names the target object. This ten- dency of English-learning 2-year-olds to "false alarm" in response to stressed novel words preceded by the article the shows that they are integrating multiple probabilistic cues to predict what kind of word is coming next. Such studies using on-line measures of children's comprehension as the spoken sentence unfolds reveal a critical dimension of emerging language competence that was impossible to monitor with precision using off-line methods. As children learn to interpret words in combination, they develop effi- ciency in integrating distributional, lexical, prosodic, and other available sources of information, enabling them to make sense of words that are known while avoiding costly interference from unfamiliar words in the sentence that are not yet known.

2. Emerging Awareness of Relations among Words

By two years of age, most children are demonstrating impressive skill at interpreting the speech that they hear around them. Several studies using preferential listening techniques (e.g., Gerken, Wilson, & Lewis, 2005; Gomez, 2002) as well as neurophysi- ological responses (e.g., Friederici, 2005) show that children in the second year are increasingly attentive to regularities in speech that are relevant to the grammatical struc- ture of the language they are learning. They have also built up a considerable repertoire of words in their production vocabularies, and are beginning to use two- or three-word combinations (e.g., mommy sock). Soon, however, utterances increase in length and com- plexity in various ways. Children add more verbs, adjectives, and other predicates to their working vocabularies, and substantively increase their use of prepositions, articles, and other closed-class forms that do grammatical work, including the productive use of inflectional morphemes (e.g., English past tense –ed). At the same time, there is also size- able variation in exactly when and how children move into more grammatically complex utterances in their everyday language use. Indeed, while some children are reported to use primarily multi-word phrases and many closed-class forms by 24 months, other chil- dren are still primarily using nouns in single-word constructions (e.g., Bates et al., 1988; Bates & Goodman, 1999). Who are the children who are more advanced in grammar at this age? Based on the norming data from the CDI: Words and Sentences, children with the highest grammar scores were also those children with the largest reported production vocabularies (r 0.85) (Bates et al., 1994). In the same data set, Marchman and Bates (1994) found that size of verb vocabulary was concurrently related to the number of reported overregular- izations of the

English past tense inflection (e.g., daddy goed), accounting for significant variance over and above chronological age.

These “mistakes” are typically viewed as a major milestone in the development of grammatical rule-based knowledge. Links between lexical development and grammar have also been reported longitudinally. Following 27 children, Bates et al. (1988) found that the best predictor of grammatical sophistication at 28 months (as measured by mean length of utterance, MLU) was size of vocabulary 10 months earlier. Bates and Goodman (1997) cite similar relationships in a sample of children followed monthly from 12 to 30 months. Other researchers have targeted children at the extremes in acquisition (e.g., late vs. early talkers), revealing that children who were delayed in early vocabulary production were later delayed in the use of grammatical forms (Paul, 1996, 1997; Rescorla & Schwartz, 1990; Rescorla, Roberts, & Dahlsgaard, 1997, 2000; Thal & Tobias, 1994; Thal & Katich, 1996; Marchman & Armstrong, 2003) and that particularly precocious children display grammar abilities that are commensurate with their vocabulary, even though they are considerably younger than peers at their same level (Thal, Bates, Zappia, & Oroz, 1996; Thal, Bates, Goodman, & Jahn-Samilo, 1997). Similar lexical–grammar links have been found in children with focal brain injury (e.g., Bates et al., 1997; Marchman, Miller, & Bates, 1991; Marchman, Wulfeck, & Saccuman, 2003; Thal et al., 1991), and Willliams syndrome (e.g., Singer-Harris, Bellugi, Bates, Jones, & Rossen, 1997).

More recently, studies have documented that lexical development and grammar are related to a similar degree in children learning more than one language, with gram- matical abilities robustly linked to lexical level in the same, but not the other, language (Marchman, Martínez-Sussmann, & Dale, 2004). Finally, strongly heritability of the relation between lexical and grammatical level has been documented in behavioral genetic studies of monozygotic and dizygotic twins (Dale, Dionne, Eley, & Plomin, 2000). In other words, even though genetic factors make a relatively weak contribution to each aspect of language assessed individually, the genetic factors that influence lexi- cal growth are the same as those that influence grammatical growth.

These studies all point to the idea that vocabulary and grammar development are highly interdependent, a view at odds with the nativist assumption that grammatical knowledge is autonomous and emerges independent of lexical knowledge. In light of the extensive individual variation that is observed in early language development, it is strik- ing that lexical and grammatical skill “hang together” so tightly over acquisition, especially when abilities that would seem to more likely to travel together are less strongly related (e.g., reported lexical comprehension and production). Such interdependence is quite natural, however, within a view of acquisition in which domain-general learning mechanisms guide the child’s construction of a working linguistic system simultaneously at many different levels, in this case, learning words and learning grammatical rules. As Bates and MacWhinney (1987) proposed many years ago, “the native speaker learns to map phrasal configurations onto propositions, using the same learning principles and representational mechanisms needed to map single words onto their meanings” (p. 163, emphasis added).

This type of domain-general continuity is directly modeled in connectionist and dynamical systems accounts of language development (e.g., Plunkett & Marchman, 1993; Elman et al., 1996; van Geert, 1998), and is at the core of probabilistic constraint-based explanations of many other psycholinguistic and developmental phe- nomena (e.g., Elman et al., 1996; Elman, Hare & McRae, 2005; Harm & Seidenberg, 2004; Tomasello, 2003). Interestingly, enhanced reliance on domain-general continuity has gained credibility in several frameworks in modern-day linguistics (e.g., Bresnan, 2001; Croft, 2001; Goldberg, 1995; Langacker, 1987). Finally, several recent studies have focused on ruling out indirect explanations for lexical–grammar links, for example, that lexical and grammatical relations derive from common influences from the environ- ment or general cognitive or linguistic intelligence (e.g., Dale et al., 2000; Dionne, Dale, Boivin, & Plomin, 2003; Marchman et al., 2004).

Clearly, there is much more to be said about early vocabulary and grammar develop- ment. Studies are continuing to map out in more and more precise ways how those domain-general mechanisms might operate (e.g., Bates & Goodman, 1999; Tomasello,2003; Elman, 2004; Naigles, 1996), and how those relations might change in character over development (Dionne et al., 2003; Tomasello, 2003). Yet, the picture that is gaining mounting empirical support portrays language acquisition as a gradual and continual process of mapping various types of linguistic entities onto communicative functions, using mechanisms that are shared across many different levels of the linguistic system.

CONCLUSION

Language is a tool to communicate, can be used to think, express feelings and through language can accept the thoughts and feelings of others. Language development begins in infancy and relies on its role in language experience, mastery and growth. The development of language skills for Early Childhood aims to enable children to communicate verbally with their environment. The context of language development includes: listening, speaking, reading and writing early. In developing children's language skills, the teacher / tutor can choose a variety of strategies and methods. Activities that can be done in developing language skills are activities that can stimulate the ability to listen, speak and write. The storytelling method is one of the methods that is widely used for Early Childhood. The news that is delivered by the teacher must attract and invite the attention of children and not be separated from the educational goals for Early Childhood.

[This page is intentionally left Blank]

CHAPTER 16
ACQUISITION OF SYNTAX AND SEMANTICS

INTRODUCTION

There are two main views about the nature of language development. These views can be traced back to the ‘nature versus nurture’ debate about how knowledge in any domain is acquired. The ‘nativist’ perspective dates back to Plato’s dialogue “The Meno”. This view emphasizes the contributions of human nature to the acquisition of knowledge. It is supposed on the nativist approach to language that children are biologically fitted, as part of the human genome, with a theory of ‘Universal Grammar’ (e.g., Chomsky 1965, 1975, 1986). Universal Grammar contains the ‘core’ principles of language, i.e., principles that are manifested in all human languages. In addition, Universal Grammar spells out particular ways in which human languages can vary; these points of variation are called parameters. Taken together, the principles and parameters of Universal Grammar establish the boundary conditions on what counts as a possible human language. Children navigate within these boundaries in the course of language development. Of course, experience determines which particular language children acquire, but nativists argue that much of the process of language acquisition is biologically driven, rather than being “data driven.” The nativist approach views language learning as the by-product of a task-specific computational mechanism, with a structure that enables children to rapidly and effortlessly acquire any human language, without formal instruction and despite considerable differences in linguistic experience. Universal linguistic principles are not learned by the computational mechanism, but are implicit in the structure of the mechanism itself – i.e., these are in the Universal Grammar. This implicit (or built in) knowledge explains how learners come to know more about language than they observe from experience. This is the nativist’s solution to ‘Plato’s Problem’.

The alternative ‘nurture’ approach views language development as the product of domain general learning mechanisms. These mechanisms embody general learning processes that are not specially tailored to acquire any particular kinds of facts about the world. Like knowledge in other domains, knowledge of language is accrued in a piecemeal fashion, based on statistical regularities in the input. According to the experience-dependent account, child language matches the input more or less, with more frequently attested constructions being mastered earlier in the course of language development. Gradually, more and more complex structures are composed, until the language of the child approximates that of an adult in the same linguistic community. Tomasello (2000) sums up this approach as follows: "When young children have something to say, they sometimes have a set expression readily available and so they simply retrieve linguistic schemas and items that they have previously mastered (in their own productions or in their comprehension of other speakers) and then "cut and paste" them together as necessary for the communication situation at hand …" (p. 77).

The nature versus nurture debate has intensified in recent years. For a few decades, linguists working within the theory of Universal Grammar pointed out the difficulty learners faced in mastering many facets of language. In the 1980's and 1990's, a great many experimental studies of children's adherence to linguistic universals were reported in the literature, leading to a picture of language development that was consistent with Universal Grammar.

Recently, there has been a shift in the opposite direction. More and more, it seems, developmental psycholinguists are exploring the possibility that linguistic facts can be learned without the kinds of abstract or implicit principles that have been proposed in the theory of Universal Grammar. Two developments have prompted this change in direction. One is the discovery that children are able to effectively learn certain linguistic properties based on statistical regularities in the input. For example, Saffran, Aslin and Newport (1996) showed that 8-month-old children could exploit statistical learning to extract information about transitional probabilities from the input. Infants inferred the existence of word boundaries between three-syllable pseudowords (nonsensical combinations of syllables). Those three-syllable sequences that crossed a word boundary were not treated by the child subjects as a 'word' during the post-test phase of the study, because there was a lower probability for such sequences to be repeated if they crossed a word boundary than if they were part of a 'word.'

The second development concerns the nature of the input available to children. The linguistic input had been assumed to be quite impoverished and, therefore, insufficient to support language learning without assistance from

Universal Grammar (Chomsky 1980). It has recently been argued, however, that the input contains relevant features in sufficient abundance to support statistically based acquisition of several seemingly complex facts about language. We will discuss this issue in the next section of the chapter.

Critics of statistical learning have pointed out limitations in statistical learning mechanisms that exploit transitional probabilities. For example, Yang (2004) showed that statistical learning mechanisms cannot reliably segment sequences of monosyllabic words, though such sequences make up the majority of the input that is directed to children. In a series of papers, Marcus (1998; 1999; Marcus, Vijayan, Rao and Vishton 1999) has shown that statistical learning mechanisms are ill suited to learning many properties of languages (also see Smith, 1996). At the same time, the arsenal of arguments and evidence in support of nativism, and against the experience-dependent ("data driven") approach to language development, has also continued to grow. Evidence in favour of the nativist perspective takes several different forms. First, experimental investigations have shown that children do not violate core linguistic principles, even in cases where they might be tempted to violate such principles if they were to adopt general-purpose learning algorithms (section 3). Second, the nativist approach is reinforced by the observation that children learn 'deep' linguistic principles that tie together apparently disparate facts about language; this is another aspect of children's linguistic competence that is not plausibly a product of experience (section 4). Third, it has been demonstrated that children know 'hidden' aspects about the meanings of certain sentences; again, it is unlikely that these aspects of meaning are learned from experience (section 5). Finally, studies have revealed that children follow the natural seams (parameters) of natural language even when child language differs from that of adults. Some features of children's nonadult linguistic behaviour, moreover, are quite unexpected on an experience-dependent account of language development (section 9).

In this chapter, we discuss these arguments and review some of the results from experimental investigations of child language. The experimental findings should be influential in the current debate about the nature of language development because, as Tomasello (2000) asserts: "Choosing between the alternative is, or should be, an empirical matter…" (p. 67). We agree. To get started, we will describe how both the experience-dependent approach and the nativist approach attempt to deal with one of the most fundamental features of language - its dependence on structure.

A. Structure Dependence

Much of the current debate in the literature focuses on the nature of linguistic operations. The example of Yes/No questions is frequently cited. At issue is the relation between declarative sentences (on the left-hand side of the arrow) and their Yes/No question counterparts (on the right-hand side). For every ordinary declarative sentence in English, there is a corresponding Yes/No question, so these structure are obviously related. But how?

(1) (a) Bill can play the sax. ⇒ Can Bill play the sax? (b) The sky is blue. ⇒ Is the sky blue?

As Chomsky (1971; 1975) observed, a simple ‘structure-independent’ hypothesis yields the correct results for much of the input that children receive. For example, the following structure-independent hypothesis will generate the Yes/No questions in (1):

Structure-Independent Rule: To form a Yes/No question, move the first verbal element {is, can, has, …} of the declarative statement to the front.

The inadequacy of this structure-independent operation is revealed when it is applied to complex examples with a modifying clause (who is beating a donkey, as in 3). The first is appears in the modifying clause; if it is moved to the front, the result is an unacceptable Yes/No question in (3).

(2) Declarative: The farmer who is beating a donkey is mean.
(3) Yes/No question: Is the farmer who _ beating a donkey is mean?

To produce the correct Yes/No question corresponding to (2), the auxiliary verb “is” following the entire subject phrase the farmer who is beating a donkey is moved to the front, yielding (4).

(4) Yes/No question: Is the farmer who is beating a donkey _ mean?

A rough formulation of the structure-dependent rule that gives the right results is something like the following:

Structure-Dependent Rule: Move the Auxiliary verb in the main clause to a sentenceinitial structural position.

Chomsky (1971) maintained that children would never adopt structureindependent hypotheses, even if the data available to children were consistent with both structure-independent and structure-dependent rules. In other words, children would not be expected to make certain kinds of mistakes in forming Yes/No questions at any stage in language development. So, for example, they are not expected to produce questions like (3): Is the farmer who beating a donkey is mean? In an elicited production study, Crain and Nakayama (1987) evoked Yes/No questions from thirty 3-5 year old children, to see if they ever made such mistakes. Although children made certain kinds of errors, they never produced questions that were consistent with structureindependent rules. (On the other hand, the kinds of nonadult responses children made were consistent with the continuity assumption; see section 9).

It has frequently been claimed by advocates of the experience-dependent approach that nativists assume that "no evidence exists that would enable a three-yearold to unlearn" mistaken structure-independent rules, if children were to initially adopt such rules (Cowie 1999; also see MacWhinney 2004, Pullum and Scholz 2002). But no reasonable nativist would endorse such a strong claim about all possible evidence. The following passage from Chomsky (1975, p.31) is often quoted as the basis of this conclusion about nativism: "A person may go through a considerable part of his life without ever facing relevant experience, but he will have no hesitation in using the structuredependent rule, even if all of experience is consistent with [the structureindependent rule]."

As this passage makes clear, Chomsky is not claiming that nobody ever has relevant experience. The issue concerns the robustness of evidence, not its existence, and children's use of the evidence, regardless of its quantity. As Lasnik and Crain (1985) note, if relevant data are not available in sufficient quantities, then at least some (and perhaps many) children won't come by them, and these children will not converge on 5 an adult grammar. But this is contrary to fact. All (normal) children converge on a system of linguistic principles that is equivalent to that of adults. Therefore, if convergence depends on there being evidence, then it must be available in abundance. Suppose to the contrary, that evidence falsifying the structure-independent hypothesis for forming Yes/No questions is not available to children before they reach their third birthday. Then many children should be observed to make structure-independent errors. But Crain and Nakayama (1987) did not find any evidence that children were adopting structure-independent rules. So either children never form structure-independent hypotheses, or there is abundant evidence available to and used by very young children. According to Cowie, "… something like the requisite guarantee can be provided when one reflects on the sheer size of the data sample to which a learner has access." (p. 219). If the requisite evidence includes sentences like (4), however, then the evidence is not readily available to children. A search of the input to English-speaking children turned up only one example of a structure like (4) out of about 3 million utterances (reported by MacWhinney 2004 using the CHILDES database; see MacWhinney and Snow 1985). Advocates of the experience-based approach have therefore proposed other sources of evidence for children. One example is (5), which is assumed to be derived from the declarative sentence represented in (6).

(5) Where's the other dolly that was in there?
(6) [the other dolly that wasAUX in there] [isAUX where]

Notice that the representation that is assigned to the declarative sentence in (6) is partitioned into a main clause [is AUX where] and a relative clause [the other dolly that was AUX in there]. To form the corresponding wh-question, (5), the main clause [is AUX where] is moved to the front (and its internal parts are inverted). Despite the absence of sentences like (7) in the input to children, if questions like (5) are readily available in the input, then these questions would be subject to a similar analysis (compare examples 6 and 8). In forming both kinds of questions, the "is AUX" in the relative clause appears first in the declarative sentence, but it remains in place, whereas the "is AUX" in the main clause is moved to the front in order to form the corresponding question.

(7) Is the farmer who is beating the donkey mean?

(8) [the farmer who isAUX beating the donkey] [isAUX mean]

As MacWhinney (2004) acknowledges, this experience-based account "requires children to pay attention to relational patterns, rather than serial order as calculated from the beginning of the sentence" (p. 891). So we should ask what distinguishes the experience-based account from the nativist account offered by Chomsky. The difference is that the experience-based account could, in principle, have learned to move the first 'is' to the front, as in the structure-independent rule described earlier. In Chomsky's view, children are incapable of any such structure-independent analyses.

Regardless of the input, according to the nativist account, children are compelled to impose a symbolic analysis onto the utterances they experience. According to the experience-based account, on the other hand, children have no such predisposition; the system children acquire depends on the statistical regularities of the input.

It remains to determine whether or not questions like (5) [Where's the other dolly that was in there?] are available in sufficient quantity in the input to children, to ensure that every child converges on a grammar that conforms to structure-dependent operations. In the present case, Legate and Yang (2002) conclude that the input does not suffice. They report the results of a search through transcriptions of the input to two young children, Nina and Adam (in the CHILDES database). The input to Nina consists of 14 critical wh-questions out of 20,651 questions overall. There were just 4 critical examples out of 8,889 questions in the input to Adam. The paucity of critical input for these children bears out Chomsky's conjecture that a child could "go through a considerable part of his life without ever facing relevant experience." Moreover, such low frequencies of relevant input make it unlikely that every child encounters the requisite evidence by the age at which they are found to adhere to structure dependence. This is a problem for the experience-based account because, without relevant input, some (perhaps many) children would be expected to commit structureindependent errors, such as (3), but this is contrary to the findings of experimental research (Crain and Nakayama 1987).

B. Avoiding errors: innate constraints versus conservatism

1. A constraint on reference.

Another distinguishing feature of the two approaches to language development is how they explain the kinds of sentences children refrain from producing, and the kinds of meanings that children do not assign to sentences. One case in point is the reference of ordinary pronouns. Notice that in the examples in (9) and (10), the pronoun he may or may not refer to the individual called the Ninja Turtle. To indicate these dual referential possibilities, we will adopt the following notation: two expressions refer to the same individual(s) only if they have the same index. So, (9) and (10) are ambiguous, because the pronoun he can have the same index as the Ninja Turtle ('1'), but one of these expressions can also be assigned an index '2' which the other expression lacks; in that case, the two expression are said to be disjoint in reference or non-coreferential.

(9) The Ninja Turtle1 danced while he1/2 ate pizza. (10) While he1 ate pizza, the Ninja Turtle1/2 danced.

Consider another sentence, (11), which also contains the pronoun he and the expression the Ninja Turtle. Unlike examples (9) and (10), (11) is unambiguous. Intuitively, the pronoun he cannot refer to the Ninja Turtle, but must refer to some other male individual. In other words, co-reference (as indicated by the assignment of the same number) is not permissible in the sentence in (11); it has the reading in (11a), but not the reading indicated in (11b).

(11) He danced while the Ninja Turtle ate pizza. (a) He1 danced while the Ninja Turtle2 ate pizza (b) *He1 danced while the Ninja Turtle1 ate pizza

There are two ways of describing the possibilities for referential interpretations of pronouns. Each of these options has been taken by one of the two approaches to language development. One way is to list the various possibilities for coreference. This is the strategy taken by the experience-dependent approach. Adopting this strategy, the list includes some way of representing the positive instances of coreference between pronouns and other expressions, so examples like (9) and (10) would be represented (somehow) in the list. Nothing would be said about the case in (11b), because this is not an instance of coreference.

The alternative strategy is to formulate a negative principle representing those cases in which coreference is prohibited, such as (11b). Nothing is said about any of the other cases, such as (9), (10) and (11a). On grounds of parsimony, Lasnik (1976) argued for the second strategy, because the list of cases where coreference is possible adds up to a huge inventory of linguistic representations, whereas a single generalization can explain mandatory non-coreference, with cases of coreference left open. Negative linguistic principles are known as constraints. So a constraint prevents coreference between pronouns and referring expressions in sentences like (11b).

Constraints are frequently invoked in arguments for nativism for the following reason. Suppose for the sake of argument that children's grammars embody constraints as negative statements; in the present example, the constraint is a prohibition against certain co-reference possibilities, as illustrated in (11b). It seems unlikely that children could 'learn' such negative facts from experience, because parental speech rarely if ever includes explicit negative evidence and learning constraints would seem to require negative evidence (see, e.g., Bowerman 1998; Brown and Hanlon 1970; Marcus 1993). Acquisition in the absence of decisive evidence is one of the main hallmarks of innate specification of knowledge.

Another hallmark of innate specification is early emergence (Crain 1991). Developmental psycholinguists have investigated the time-course of the acquisition of constraints in pursuit of the early emergence hallmark of innateness. Of course, even innate principles need not emerge early in the course of development. Just as some properties of physical development are biologically timed to appear long after birth (e.g., a second set of teeth), certain aspects of linguistic knowledge might become operative only at a certain maturational stage of development (Borer & Wexler 1987). But the earlier complex principles emerge in child grammars, the more difficult it would be for the experience-dependent approach to explain the facts, because early emergence of knowledge compresses the evidential basis for learning.

The question of children's knowledge of the constraint on coreference was pursued in a comprehension experiment by Crain and McKee (1985). In this experiment, children encountered sentences like (12) in circumstances appropriate to both interpretations. On one interpretation of (12), the pronoun he and the referring expression the Ninja Turtle have the same index. This is called the backwards anaphora interpretation. It is ‘backwards’ in the sense that the pronoun comes first. More typically a pronoun follows the expression with which it is anaphorically linked. On the alternative interpretation of (12), the Ninja Turtle is not co-indexed with the pronoun. The pronoun is said to have extrasentential reference on this reading; it refers to an individual who is not mentioned in the sentence. (12) While he ate pizza, the Ninja Turtle1/2 danced.

The experimental procedure used in the Crain and McKee (1985) study was the TruthValue Judgment Task. As the name suggests, this procedure requires subjects to judge the truth or falsity of a sentence, according to its fit to the context. Two experimenters are needed to conduct the Truth Value Judgment Task. One experimenter uses toys and props to act out a situation corresponding to one interpretation of the target sentence. A second experimenter manipulates a puppet; we often use Kermit the Frog as the puppet. Following each situation, Kermit the Frog says what he thought happened on that trial. When Kermit the Frog accurately describes something that happened in the story, the child is instructed to reward him, say with a strawberry. Sometimes Kermit doesn't pay close attention, however, and he says the wrong thing. In that case, the child is instructed to give Kermit something to remind him to pay closer attention, say a rag. These procedures make it fun for children to attend to Kermit’s statements. Without the rag ploy children are reluctant to say that Kermit has said anything wrong. Notice that both (a) the events corresponding to the meaning of the target sentence, and (b) the target sentence itself are provided for the children. This allows unparalleled experimental control and at the same time reduces extraneous processing demands that are present in comprehension tasks in which children are required to act out events themselves. For ambiguous sentences such as (12), the same sentence was presented on two separate occasions, in two contexts. In one context for (12), the Ninja Turtle was dancing and eating pizza; in the other, someone else ate pizza while the Ninja Turtle was dancing. Kermit uttered the same sentence following both situations. The results were that children accepted the backward anaphora reading about two-thirds of the time, in appropriate contexts. The extrasentential reading of the pronoun was accepted only slightly more often. Only one of the 62 two- to five-year-old children (mean age 4;2) interviewed in the Crain and McKee study consistently rejected the backward anaphora interpretation. To test children's knowledge of the constraint against coreference, there was another condition in the experiment. In this condition, sentences like (13) were presented in situations corresponding to the meaning that is ruled out by the noncoreference constraint. For (13), the situation was one in which the Ninja Turtle danced and ate pizza at the same time. If children adhered to the constraint prohibiting coreference, they were expected to reject (13) as an accurate description of this situation. (13) He danced while the Ninja Turtle 2 ate pizza. In fact, the child subjects judged sentences like (13) to be false almost 90% of the time. In the context for sentence (13), it was clear that some other salient (male) character did not dance while the Ninja Turtle ate pizza. This made it reasonable for children to give a "No" response, provided that their grammars made a "Yes" response inappropriate. This characteristic of the task is called plausible dissent. The need for plausible dissent in experiments is discussed in detail in Crain and Thornton (1998). The findings show that even 2- and 3-year olds prohibit backwards anaphora only when structural conditions (involving c-command) dictate that they should. It is important to appreciate that this experiment provides further evidence that children do not rely on their linguistic experience in making judgements about the appropriate mappings of sentences with their meanings. Since there is nothing in children's experience to tell them which sentence/meaning pairs are NOT allowed, there is no way to learn the structural constraint prohibiting coreference if this ‘negative statement’ is part of children’s grammars. It has been proposed that that the same constraint that prohibits coreference in sentences like (14) also govern coreference relations in some discourse contexts. Consider the short discourses shown in (15) and (16).

(14) *He sent the letter to Chuckie’s1 house.

(15) Speaker A: I know where he sent the letter. Speaker B: Me too. To Chuckie’s2 house

(16) Speaker A: I know where he sent the letter. Speaker B: Me too. *To Chuckie’s1 house

It is intuitively clear that in (14) the pronoun he cannot refer to Chuckie, but must refer to some other salient male in the conversational context. The judgments about coreference, then, are similar to the judgments for sentences like (13). This raises the possibility that the same constraint governs both linguistic phenomena. A recent proposal to this effect was made by Merchant (2005) (cf. Hankamer 1979; Morgan 1973, 1989). The idea is that part of the structure of the statement by Speaker A is reconstructed by Speaker B, but subsequently deleted. This is illustrated in (17).

(17) Speaker A: I know where he sent the letter. Speaker B: Me too. He sent the letter to Chuckie's2 house.

Even though only a fragment answer (To Chuckie's house) is actually produced by Speaker B, it has the same propositional content as a full sentence.

In a recent study, Conroy and Thornton (2005) presented both full sentences and discourse sequences to twenty English-speaking children (mean age 4;6), to see whether children made similar judgments in response to both complete sentences and discourse sequences. On a typical trial of the relevant experimental condition, one of the characters, Tommy, was preparing to send a letter to Chuckie's house, but then decided against it. In response, Chuckie sent a letter to his own house. Against this backdrop, half of the time, children heard a complete sentence (He sent the letter to Chuckie's house), and half of the time, they heard a discourse like the one in (17). The main finding was that children rejected both the full sentences and the discourse sequences an equal proportion of the time (86% rejections, as compared to 89% rejections, respectively). This finding is consistent with the hypothesis that the same negative constraint against coreference underlies children's responses to both phenomena. On the experience-dependent approach, there is no reason to expect children's responses to both phenomena to coincide, but there is nothing in the approach that rules out this possibility either.

2. A constraint on contraction.

Another example of a constraint governs where contraction may and may not occur. In English, this constraint prevents the verbal elements want and to to be contracted to form wanna in certain kinds of sentences, although wanna-contraction is permitted most of the time. Examples (18)-(21) illustrate permissible contractions. Example (22a) illustrates an impermissible contraction.

(18) (a) Who does Arnold wanna make breakfast for? (b) Who does Arnold want to make breakfast for?

(19) (a) Does Arnold wanna make breakfast for Maria? (b) Does Arnold want to make breakfast for Maria?

(20) (a) Why does Arnold wanna make breakfast? (b) Why does Arnold want to make breakfast?

(21) (a) I don't wanna make breakfast for Arnold or Maria. (b) I don't want to make breakfast for Arnold or Maria.

(22) (a) *Who does Arnold wanna make breakfast? (b) Who does Arnold want to make breakfast?

All of the questions in these example begin with wh-words (who, what, why, where, even how) and will be called wh-questions. According to a standard account of wanna contraction, wh-questions are formed by movement of a wh-phrase from one position at an underlying level of representation to another position, on the surface, where it is pronounced. A further assumption of the account is that a record, which we abbreviate as t (for 'trace'), is left behind at the site of the origin of the wh-movement. In (23) the wh-phrase originates in the subject position of the embedded infinitival clause want t to kiss Bill. When the wh-phrase starts out between want and to, as in (16), the trace left behind by wh-movement blocks the contraction of want and to. This explains why (23b) is ruled out. The same account explains the unacceptability of (22a).

(23) (a) Who do you want t to kiss Bill? Subject Extraction (b) *Who do you wanna kiss Bill?

By contrast, in (24), the formation of the wh-question requires the movement of the whphrase from the object position of the embedded infinitival clause. In that case, the trace does not intervene between want and to, so wanna-contraction is permitted.

(24) (a) Who do you want to kiss t ? Object Extraction (b) Who do you wanna kiss?

These facts invite the following generalization: Contraction of the verbal elements want and to is blocked if the trace of wh-movement intervenes between them. In declaratives, the constraint on contraction is irrelevant, so contraction is tolerated. As examples (18)-(21) indicate, much of the evidence available to children learning English runs counter to the constraint. That is, contraction of want and to is licensed in general -- (22) is an exception to the rule. If the grammars of English speaking children lacked the constraint on contraction of want and to (across the trace of a moved wh-word, then child English would include more sentences than adult English does. In other words, without the constraint, children would over generate, and would produce sentences like (22a) and (24b) with illicit contraction of want and to. Children who lack the constraint on contraction across a trace should permit contraction to a similar extent in both subject and object extraction questions. To test children's adherence to the constraint, an experiment was designed to elicit relevant questions from children (Thornton 1990, 1996).

This permitted a comparison of the proportion of contraction by children in questions like (24) with contraction in questions like (23). The finding was that the 21 children interviewed (mean age 4;3) produced contracted forms more than half the time (57%) in questions like (23), but the same children produced contracted forms less than 10% of the time in questions like (24), where contraction is outlawed by the constraint. The linguistic constraint that prohibits wanna-contraction also applies to a variety of other constructions, but not in ways that can easily be determined on the basis of the primary linguistic data. For example, the constraint prohibiting contraction across a trace governs a linguistic phenomenon known as is-contraction. A good case can be made that is contracts to its right, despite the orthographic convention which links an 's with the word to its left. The paradigm in (25) shows that is can contract when there is no trace to its immediate right, as in (25b), but contraction is blocked when there is a trace to its immediate right, as in (25d).2

(25) (a) Do you know what that is doing t up there? (b) Do you know what that's doing t up there? (c) Do you know what that is t up there? (d) *Do you know what that's t up there?

Having witnessed two applications of the constraint on contraction, it is important to ask how a learning-theoretic account could explain the generalization that relates wanna-contraction and is-contraction. The constraint applies to linguistic phenomena that bear little superficial resemblance. In the wanna-contraction paradigm, the constraint prevents contraction across the subject position of an embedded infinitival clause, whereas in the is-contraction paradigm, the same constraint prevents contraction across the object position in a tensed clause. Until a wide range of linguistic phenomena was considered, including both positive and negative data, linguists failed to see that the two phenomena were related. Assuming that language-learners do not have access to such complex arrays of positive and negative data, nativists conclude that language-learners must have an advantage over linguists, in knowing the linguistic constraint in advance of encountering the limited primary linguistic data to which they have access. Returning to child language, the nativist is compelled to predict that children will adhere to the constraint on contraction across a trace in both constructions. Another twelve 2- to 4-year-old children (mean age 3;8) participated in an elicited production experiment designed to assess their knowledge of the constraint that prohibits iscontraction. The finding was the complete absence of illicit productions. Illicit contraction is apparently prevented by the constraint. These mutually supporting findings suggest that the same constraint rules out (24b) and (25d).

Experience-based accounts of language acquisition take a different stance on the acquisition of ‘constraints’. Their approach is to list the positive cases. By supposing that children are ‘conservative’ learners, in the sense that their grammars are directly tied to experience, such accounts avoid the problem of learning negative constraints in the absence of negative evidence. Being conservative, learners never produce forms they do not encounter. For example, Pullum and Scholz (2002; p 16) use the conservative learning strategy to explain how the linguistic expressions that co-occur with (= complements) subordinate words like than and that are learned, as summarized in (26).

(26) They wanted more than was available. a) … ‘more than + finite VP’ b) * … ‘more that + finite VP’ "If the types of constituents that can occur as complements to subordinating words like that and than are learned piecemeal from positive examples, then the pattern 'more than + finite VP' will be learned (after examples like They wanted more than was available), but 'V + that + finite VP' will not be learned, because no examples of that sort will ever be encountered."

As this quote makes clear, experience-based accounts are committed to piecemeal acquisition from positive example, so that the absence of generalizations beyond the input are explained. As Cowie (1999, p. 223) remarks:
“the non appearance of a string in the primary linguistic data can legitimately be taken as constituting negative evidence.”

But, as Crain and Pietroski (2001) point out, the conservative learner will have to keep detailed records of all kinds of grammatical distinctions in order to avoid potential pitfalls in sentence production. For example, the declarative in (27) has the wh-question counterpart in (28) (where someone has been turned into who, and moved to the front of the sentence). But, the subordinate word that must not appear in the wh-question corresponding to (28), as the unacceptability of (29b) illustrates.

(27) He is hoping that someone is coming to visit.

(28) [who is he hoping] [that _ is coming to visit]

(29) a) Who is he hoping is coming to visit? b) *Who is he hoping that is coming to visit?

Consider next how children would represent the absence of wanna-contraction in the kinds of wh-questions discussed in the last section. Children would need to distinguish between word strings that differ in the nature of the wh-phrase because, as we saw, wanna contraction is permitted in 'why' questions, but not in all 'who' or 'what' questions. In 'who' and 'what' questions, wanna-contraction is permitted if the verb in the clause following wanna is transitive, but only if the trace of the moved wh-phrase follows the verb, rather than precedes it (cf. examples 23 and 24). In short, children must encode the distinction between subject and object position, as well as between transitive and intransitive verbs. In the simplest case, children would require a statistical learning mechanism that operates on labeled strings that are six words long (wh-phrase, auxiliary verb, subject NP, want, to, verb). To further reinforce the need for detailed record-keeping on the experience dependent account, let us look at the distribution of the expression at all.

Other words with similar distributional patterns are any, much, and ever – the class of such expressions is referred to as negative polarity items. Example (30) illustrates that the universal quantifier every licenses the negative polarity item at all in the subject phrase (e.g., every politician who favors the rich or every politician in this room), but negative polarity items are not permitted in the predicate phrase of such sentences. Suppose a learner who encountered (30a) formed the broader generalization that every licenses negative polarity items in either position. Such a learner would overgenerate, i.e., she would produce the unacceptable (30b) as well as the acceptable (30a). Conservative learning is necessary to hold over generation in check. (30) a) Every politician who favours the rich at all is in this room. b) * Every politician in this room favours the rich at all.

Example (31) reveals that both ‘local’ and ‘distant’ negation (not or n’t) license negative polarity items. And (32) reveals that some linguistic contexts with negation license negative polarity items, but others do not. (31) a) Bush doesn’t believe that liberals favour the poor at all. b) Bush believes that liberals don’t favour the poor at all.

(32) a) The news that Bush won didn’t surprise the Supreme Court at all.
b) * The news that Bush didn’t win surprised the Supreme Court at all.

If children are to avoid the kinds of overgeneration illustrated in the (b) examples of (29), (30) and (32), then they must keep track of all of the relevant distinctions in the linguistic contexts that license negative polarity items, and ones that do not. This prediction of the experience-based account seems highly implausible to the nativist, because the relevant distinctions that children would need to keep track of are so subtle and so numerous (see Crain and Pietroski 2001, p. 172-173). The kinds of record keeping that is needed to mimic linguistic constraints would seem to be beyond the capacity of certain statistical learning mechanisms, such as connectionist or parallel distributed processing networks. These networks rely on local regularities -- i.e., changes in the "connection between one unit and another on the basis of information that is locally available to the connection" (Rumelhart and McClelland 1986, p 214). According to Rumelhart and McClelland, such models "provide very simple mechanisms for extracting information from an ensemble of inputs without the aid of sophisticated generalizations or ruleformulating mechanisms." Such models are evidently incapable of learning the kinds of linguistic facts that children learn, such as facts about the ‘displacement’ of wh-phrases, the consequences of wh-movement for contraction, and the permissible locations of negative polarity items.

3. An Unexpected Generalization

There are more arrows in the nativist’s quiver. Another reason for questioning the experience-based account of language acquisition is the lack of explanations for (a) the generalizations formed by children in the course of language development, (b) crosslinguistic generalizations. As in any other science, progress is made in linguistics when apparently unrelated facts can be amalgamated. As the physicist Richard Feynman (1963, p 23-24) remarks: “The things with which we concern ourselves in science appear in myriad forms, and with a multitude of attributes. … Curiosity demands that we ask questions, that we try to put things together and try to understand this multitude of aspects as perhaps resulting from the action of a relatively small number of elemental things and forces acting in an infinite variety of combinations. … In this way we try gradually to analyze all things, to put together things which at first sight look different, with the hope that we may be able to reduce the number of different things and thereby understand them better.”

In the previous section we observed that negative polarity items like at all, much, ever and any, are licensed in certain linguistic contexts but not in others. For example, we saw that at all could appear in the subject phrase of the universal quantifier, but not in the predicate phrase. This and other asymmetries are illustrated in (33)-(35).

(33) a) Every linguist who agreed with any philosopher in this room.
b) * Every linguist in this room agreed with any philosopher.

(34) a) If any linguist enters the gym, then Geoff leaves.
b) * If Geoff leaves, then any linguist enters the gym.

(35) a) Geoff went to the gym before any linguist.
b) * Geoff went to the gym after any linguist.

The point we made earlier about such asymmetries was that, on a data-driven approach, children could avoid producing the unacceptable (b) examples only by keeping careful records of the actually occurring expressions and (re)producing only the examples that were attested in the input. Otherwise, illicit examples could be generated.

There is another aspect of these asymmetries that we would draw to your attention. Another seemingly quite distinct phenomenon is manifested in the same linguistic contexts that permit any. The second phenomenon is the interpretation of disjunction (or in English). Although these linguistic phenomena are radically different in character, the fact that they are manifested in the same linguistic contexts argues that they should be amalgamated. It is worth asking how the alternative approaches to language development can achieve the amalgamation.

First it will be helpful to describe the interpretation of disjunction in logic and in natural language. In classical logic, the logical expression for disjunction '$\vee$' is assigned truth conditions associated with inclusive-or. This means that formulas of the form $A \vee B$ are true when A is true (but not B), when B is true (but not A), and when both A and B are true. Such a formula will be FALSE only when both of its disjuncts are false. If the original disjunction is negated, it will have the opposite truth conditions. So $\neg (A \vee B)$ will be TRUE only when both of its disjuncts are false. Because disjunction is assigned the truth conditions of inclusive-or, a formula in which disjunction appears in the scope of negation $\neg (A \vee B)$ will be true if and only if both A and B are false. Let us refer to these truth-conditions of disjunction under negation as the ‘neither … nor …’ interpretation. The ‘neither … nor …’ interpretation can also be rendered as a conjunction, where both of the disjuncts from $\neg$ (A $\vee$ B) and negated. This is stated in one of De Morgan's laws, where the symbol ‘$\Rightarrow$’ indicates logical entailment, and the symbol ‘$\wedge$’ represents conjunction (English and).

$\neg(A \vee B) \;\Rightarrow\; \neg A \wedge \neg B$

This logical entailment will be referred to the 'conjunctive' entailment of disjunction in the scope of negation. To some degree, natural language mirrors classical logic. For example, when disjunction appears in simple negative sentences in English, the interpretation is consistent with De Morgan’s law. Consider the sentence in (36). Adult speakers of English interpret (36) to entail (37).

(36) John doesn’t speak French or Spanish.

(37) John doesn’t speak French and John doesn’t speak Spanish.

In short, in English disjunction under negation yields the kind of conjunctive entailment described in De Morgan’s law. It follows that the word for disjunction in English (or) has the truth-conditions associated with inclusive-or in classical logic. De Morgan’s law is just the tip of the iceberg, however. Disjunctive statements yield conjunctive entailments in a great many other linguistic contexts as well as in simple negative sentences (Chierchia 2004). Example (38) shows that sentences with the universal quantifier every generate conjunctive entailments. Similarly, if disjunction is in the antecedent of a conditional, the result is logically equivalent to a conjunctive statement, as (39) illustrates. And (40) reveals that the preposition before yields conjunctive entailments.

(38) a) Every student who speaks French or Spanish is in this room. b) $\Rightarrow$ every student who speaks French is in this room and every student who speaks Spanish is in this room

(39) a) If Ted or Kyle enters the gym, then Geoff leaves.

b) ⇒ if Ted enters the gym, then Geoff leaves and if Kyle enters the gym, then Geoff leaves

(40) a) Geoff went to the gym before Ted or Kyle.
b) ⇒ Geoff went to the gym after Ted and Geoff went to the gym before Kyle

We have already witnessed these linguistic contexts, in the discussion of the contexts that license negative polarity items (repeated here).

(41) Every linguist who agreed with any philosopher is in this room.
(42) If any linguist enters the gym, then Geoff leaves.
(43) Geoff went to the gym before any linguist.

As these examples indicate, conjunctive entailments of disjunction are generated in precisely the same linguistic contexts in which any is permitted. Moreover, conjunctive entailments are not enforced when disjunction is in the predicate phrase of the universal quantifier every, or when it is in the consequent clause of conditionals, or when it follows the preposition after. In short, wherever negative polarity items are not licensed, conjunctive entailments of disjunction are not generated. To illustrate, there is no conjunctive entailment in (44). To see this, notice that (a) and (b) are not contradictory, which would be the case if (44a) made a conjunctive entailment.

(44) a) Every student in this room speaks French or Spanish. b) every student in this room speaks French or Spanish, but no one speaks both c) * ⇒ every student in this room speaks French and every student in this room speaks Spanish

With these parallels in mind, we can assert the following descriptive generalization.

(45) All (and only) those linguistics contexts that license any yield conjunctive entailments of disjunction.

The linguistic generalization in (45) represents a challenge to the experiencebased approach, because there is no apparent mechanism that would enable children to learn the phenomena under consideration. First consider the asymmetry in the acceptability of (46a) as compared to (46b).

(46) a) * Every student read anything. b) Every student who read anything passed the exam.

We have seen that a ‘conservative’ learner could avoid producing (46a) by keeping track of when negative polarity items like any occur, and using these items only in constructions where they have been attested in the input. In addition to the need for detailed records of attested constructions, if language learners are conservative in this way, admitting into their grammars only principles that generate expressions encountered in the linguistic environment, then there is a danger that they will undergenerate, such that their grammars will be weaker than adult grammars. Such learners would not achieve a state of linguistic competence that allows for production and comprehension of sentences never encountered. As Pinker (1990, p. 6) remarks, "… children cannot simply stick with the exact sentences they hear, because they must generalize to the infinite language of their community."

Look next at the asymmetry in (47). Here, no particular word is at issue. The word or is permitted in both sentences. The relevant distinction is in the interpretation of or: in (47a) or is assigned a ‘not both’ interpretation; in (47b) it makes a conjunctive entailment. The relevant distinction in this case cannot be based on any distributional analysis of the occurrence or non-occurrence of a particular (kind of) word; the distinction that must be drawn by the learner concerns the different interpretations that the same word (or) receives when it appears in different linguistic environments.

(47) a) Every student wrote a paper or made a classroom presentation b) Every student who wrote a paper or made a classroom presentation passed the exam

It is unclear how children figure out this interpretive distinction on the experiencedependent account. Worse yet is the fact that the linguistic contexts that permit any and license conjunctive entailments are highly correlated. It is implausible, to say the least, that English-speaking children are somehow informed by their caretakers that the same linguistic contexts that license any also give rise to conjunctive entailments for or. Without triggering evidence, children may require assistance from principles that operate at the ‘core’ of natural language, rather than on the surface.

The alternative to learning is innate specification. According to Universal Grammar, there is a common property that governs the insertion of negative polarity items, and the licensing of conjunctive entailments of disjunction. The common property is called downward entailment. A linguistic expression is downward entailing if it generates valid inferences from general claims about things, to specific claims about those things.3 The examples in (48) demonstrate that all of the linguistic expressions under consideration have the defining property of downward entailment, since it is valid to substitute claims about sets of things (being a Romance language) with claims about subsets of those things (French, Spanish, Italian …). (48)

a) Every student who speaks a Romance language likes to travel ⇒ every student who speaks French likes to travel

b) If a student speaks a Romance language, she likes to travel ⇒ if a student speaks French, she likes to travel.

c) John went to Europe before learning a Romance language. ⇒ John went to Europe before learning French.

So, downward entailing expressions have three properties. They license downward entailments (i.e., inferences from general statements to specific statements), they create conjunctive entailments when they are combined with disjunction, and they license negative polarity items like any and at all. This provides us with empirical tests to assess children’s knowledge of downward entailing operators. The pattern of inference is in a 'downward' direction due to the presence of these linguistic expressions. In the absence of a downward entailing expression, the pattern of inference is typically 'upward' such that inferences are sanctioned from claims about sets of things to claims about supersets of those things. So, for example, Mary bought a Ferrari entails Mary bought a car, but not vice versa.

4. Downward entailment in child language

There have been a number of experimental studies on English-speaking children's interpretation of disjunction in the scope of negation (Chierchia et al. 2001; Gualmini et al. 2001; Gualmini and Crain 2002; Crain et al. 2002; Gualmini and Crain 2004). These studies have revealed that 4- to 5-year-old English-speaking children are aware of the conjunctive entailment of disjunction under negation. A representative example is an experiment by Crain et al. (2002) using the Truth Value Judgment Task (see Crain and Thornton 1998 for extensive discussion of this task). On a typical trial of the experiment, sentence (49) was produced by a (wizard) puppet as a prediction about how events would unfold in a story.

(49) The girl who stayed up late will not get a dime or a jewel.

It subsequently turned out that the girl who stayed up late received a jewel, but not a dime. English-speaking children (mean age 5;0) correctly rejected sentences like (49) 92% of the time in experimental contexts such as this. Children's stated reason for rejecting (49) was that the girl who stayed up late had received a jewel. It is evidence that, in children's grammars, (49) entails that the girl would receive neither a dime nor a jewel. This is logically equivalent to the conjunction: the girl would not receive a dime and she would not receive a jewel. This conjunctive entailment follows from (49) only if the disjunction operator or is assigned the inclusive-or interpretation, as in classical logic. The fact that children interpreted or as inclusive-or is difficult to explain on the experience-dependent approach, because children have little direct evidence for the inclusive-or interpretation of disjunction. The majority of the input to children consists of positive statements. In most positive statements, the use of or does not conform to classical logic. Instead, the use of or implies exclusivity (the 'not both' reading) although it does not entail it.

The implicature of 'exclusivity' for or stems from the availability of another statement, with and, which is more informative (if both statements are true). The statement with and is more informative because a statement of the form A and B is true in a subset of circumstances that verify a statement of the form A or B. For this reason, the expressions or and and form a scale based on information strength, with and being more informative (stronger) than or in contexts that verify sentences with either expression. A pragmatic principle

Be Cooperative (cf. Grice 1975) entreats speakers to be as informative as possible. Upon hearing someone use the weaker term on the scale or, listeners infer that the speaker, who was being cooperative, was not in position to use the stronger term and. So the speaker is taken to imply the negation of the sentence with the stronger term: this yields the derived meaning: A or B, but not both A and B. Adult use of or is clearly governed by this scalar implicature; adults avoid using A or B in situations in which both A and B are true. Consequently, the vast majority of children's experience is consistent with the conclusion that natural language disjunction is exclusive-or, and not inclusive-or (see Crain, Goro and Thornton, in press). It should come as no surprise, then, that the same children accepted sentences like (50) 87% of the time in a context in which the girl who stayed up late had received a jewel, but not a dime. This is the same context that resulted in children's rejection of (50).

(50) The girl who didn't go to bed will get a dime or a jewel.

It ought to come as a surprise for the experience-based approach that children interpret natural language disjunction in accordance with classical logic despite the paucity of evidence for this interpretation in the input. The evidential 'gap' is even more extreme in languages like Japanese, Hungarian, Chinese, and so on, where the adult use of disjunction violates De Morgan's law even in simple negative sentences (see section 9.2). Nevertheless, cross-linguistic studies of Japanese-speaking children and Chinese-speaking children demonstrate their steadfast adherence to De Morgan's law for the conjunctive entailment of disjunction in simple negative sentences, which is possible only if these children interpret disjunction as inclusive-or Children's knowledge of the asymmetry involving the universal quantifier every has also been demonstrated in the literature. Before we discuss the findings, it will be useful to clarify a few more semantic properties of the universal quantifier. The universal quantifier is a Determiner, like no, some, both, the, three, and so forth. Structurally, Determiners combine with a noun (student) or a noun phrase (student in this room) to form a grammatical unit – like every student or every student in this room. The noun (phrase) that every combines with what is called its Restrictor (abbreviated by the subscript 'R' in the schema in 51). Once every combines with its Restrictor, the entire unit can then be combined with a predicate phrase (e.g., swims, speaks French or Spanish, etc.). The predicate phrase is called the Scope of the universal quantifier (abbreviated ‘S’ in 51). If the disjunction operator or appears in the Scope of every, it has an ‘exclusive-or’ (‘not both’) interpretation. For example, the sentence John speaks French or Spanish implies that John speaks French OR John Speaks Spanish, but not both. As before, we attribute the ‘not both’ reading to a semantic/pragmatic implicature; this interpretation is not taken as evidence of an ambiguity in the meaning of or in English, or in any natural language.

(51) Every R[…..……..…] S[…. or …….] = exclusive-or Every R[… or …] S[…………...] = conjunctive entailment

Example (52) shows that the negative polarity item any is permitted in the Restrictor of the universal quantifier every, but not in its Scope. This illustrates the descriptive generalization that any may only appear in linguistic contexts that license the conjunctive entailment of disjunction.

(52) Every R[… any …] S[…*any …]

Several studies have investigated the truth conditions children associate with disjunction in the Restrictor and in the Scope of the universal quantifier (e.g., Boster and Crain 1994; Gualmini, Meroni and Crain 2003). Using the Truth Value Judgment
21
methodology, children were asked in these studies to evaluate sentences like those in (53) and (54), produced by a puppet, Kermit the Frog.

(53) Every woman bought eggs or bananas.

(54) Every woman who bought eggs or bananas got a basket.

Sentences like (53) were presented to children in a context in which some of the women bought eggs, but none of them bought bananas. The child subjects consistently accepted test sentences like (53) in this condition, showing that they assigned an exclusive-or interpretation to disjunction in the Scope of the universal quantifier, every. Children were presented with sentences like (54) in a context in which women who bought eggs received a basket, but not women who bought bananas. The child subjects consistently rejected the test sentences in this context. This finding is taken as evidence that children generated a conjunctive entailment of disjunction in the Restrictor of every. This asymmetry in children's responses demonstrates their knowledge of the asymmetry in the two grammatical structures associated with the universal quantifier – the Restrictor and the Scope. Taken together, the findings are compelling evidence that children know that the Restrictor of every is downward entailing, but not its Scope. There are only a handful of studies bearing on the development of polarity sensitivity in children, but what little is known is consistent with the conclusion that young children produce and avoid negative polarity items in the same linguistic contexts as adults do (O'Leary and Crain 1994: Thornton 1995; van der Wal 1996). An experiment by O'Leary and Crain is representative. These researchers used a Truth Value Judgment task with an elicitation component. In the task, the puppet, Kermit the Frog, often produced false descriptions of the events that had taken place in the story. Whenever Kermit the Frog failed to accurately state what had happened in a story, children were asked to say 'what really happened.' The experimenter who was manipulating Kermit produced sentences like those in (55) and (56).

(55) Kermit: Every dinosaur found something to write with. Child: No, this one didn't find anything to write with.

(56) Kermit: Only one of the reindeer found anything to eat. Child: No, every reindeer found something to eat.

In the condition illustrated by (55), Kermit's statement had a universal quantifier every, which does not tolerate negative polarity items, such as anything, in its Scope; instead, the (positive polarity) expression something was used. Eleven children (mean age 4;10) participated in the study. These children's responses frequently contained the negative polarity item anything in linguistic contexts that license it. In another condition, illustrated in (56), Kermit's statement contained the negative polarity item anything. However, in correcting Kermit, children consistently used the universal quantifier every, so the linguistic context forced children to avoid repeating the same item.

These findings make it clear that children have mastered some, if not all, of the requisite knowledge of downward entailment, which underlies the appropriate use and avoidance of negative polarity items.

These findings are a challenge to the experience-dependent approach. On that approach, as we noted earlier, it is conceivable that children could master the facts about the distribution of negative polarity items, such as any, based on statistical properties of the input. Children would have to be exceedingly accomplished at keeping track of the linguistic environments that license such items, however, to avoid producing them in illicit environments. The challenge posed by the asymmetry in the interpretation of disjunction or, in the Restrictor versus the Scope of the universal quantifier every is more formidable, since the distinction is one of interpretation and does not involve the distribution of lexical material.

5. An abstract structural property: c-command

In addition to the 'core' semantic notion, downward entailment, a structural property cuts across all of the phenomena we have been discussing. This structural property is known as c-command.4 For example, in order for disjunctive statements to license conjunctive entailments, the downward entailing expression must c-command the disjunction operator. Moreover, in order to license negative polarity items, the downward entailing expression must c-command the position where the item is introduced. In the (a) examples in (57) and (58), negation (n't) c-commands at all, and or, respectively; c-command is absent in the (b) examples. This explains why (57b) receives a 'not both' interpretation, rather than licensing a conjunctive entailment, as in (58a).

(57) a) The news that Bush won didn't surprise the Supreme Court at all.
b) * The news that Bush didn't win surprised the Supreme Court at all.

(58) a) The news that Bush won didn't surprise Karl or Jeb. b) The news that Bush didn't win surprised Karl or Jeb

The structural relation c-command extends to many other linguistic phenomena. For example, c-command governs the constraint on coreference that prevents the ordinary pronoun he from picking out the same individual as the expression the Ninja Turtle in (58) (previous example 8). And another constraint invoking c-command explains why the reflexive pronoun himself must be coreferential with the father, and not with the Ninja Turtle or Grover, in (60).

(59) He was dancing while the Ninja Turtle ate pizza.

(60) Grover said that the father of the Ninja Turtle fed himself.

An expression A c-commands another expression B in a phrase structure diagram if there is a path that extends above A to the first branching node, and then proceeds down to B. As these examples illustrate, the structural notion of c-command runs through the principles of Universal Grammar. This structural notion would have to emerge from the statistical regularities of the input on the experience-dependent account of language development, but this too seems highly implausible, since c-command governs a wide array of phenomena that concern meaning, and not just the form of sequences of natural language expressions.

6. Universal linguistic principles

Another challenge for the experience-based account of language development is to explain why many of the phenomena we have been discussing appear in languages other than English (as well as in English). This is confirming evidence that the principles underlying these phenomena run deep beneath the surface. For example, the conjunctive entailments of disjunction under negation, and in sentences with the universal quantifier, are manifested in Chinese (61) and in Japanese (62), just as in English. It is likely that all languages exhibit the same linguistic behaviour. It may turn out that natural language disjunction is always inclusive-or, as in classical logic, and so the following generalization may be advanced: universally, disjunction yields a conjunctive entailment when it is c-commanded by a downward entailing operator.

Conjunctive entailments for disjunction (huozhe) in Chinese (61) a) Mali meiyou shuo-guo Yuehan hui shuo fayu huozhe xibanyayu. Mary not say-Perf John can speak French or Spanish "Mary didn't say that John spoke French or Spanish" ⇒ Mary didn't say that John spoke French and Mary didn't say that John spoke Spanish

b) Meige [hui shuo fayu huozhe xibanyayu de] xuesheng dou tongguo-le kaoshi every can speak French or Spanish DE student DOU pass-Perf exam "Every student who speaks French or Spanish passed the exam" ⇒ every student who speaks French passed the exam and every student who speaks Spanish passed the exam

Conjunctive entailments for disjunction (ka) in Japanese (62) a) Mary-wa [French ka Spanish-wo hanas-u] gakusei-wo mi-nakat-ta Mary-TOP or -ACC speak-pres student-ACC see-neg-past "Mary didn't see a student who speaks French or Spanish" ⇒ Mary didn't see a student who speaks French and Mary didn't see a student who speaks Spanish

b) [French ka Spanish-wo hanasu] dono gakusei-mo goukakushi-ta or -ACC speak every student pass-exam-past "Every student who speaks Icelandic or Swahili passed the exam" ⇒ every student who speaks Icelandic passed the exam and every student who speaks Swahili passed the exam.

A common rejoinder by advocates of the experience-dependent approach is that linguistic universals are part of (an innately specified) logic, and are not specific contingent properties of natural language. For example, Goldberg (2003) contends that "cross-linguistic generalizations are explained by appeal to general cognitive constraints together with the functions of the constructions involved." Attributing cross-linguistic generalizations to general cognitive constraints, rather than to specific linguistic constraints, may be plausible in certain cases, such as the basic interpretation of disjunction as inclusive-or. However, other features of natural language resist such an explanation. We will mention just two counterexamples here.

First, we have already seen that the interpretation of disjunction across languages is complicated by properties specific to natural language and not manifested in systems of logic. Recall that most positive declarative sentences with disjunction carry an implicature of exclusivity (i.e., the 'not both' reading of disjunction). This scalar implicature appears to be a universal property of natural languages, though the computation of such implicatures emerges late in the course of language development (see Chierchia et al. 2001; Guasti et al. 2005). Setting aside such complications, the basic interpretation of disjunction can perhaps be chalked up to ‘general cognitive constraints.’ In any event, this implicature is cancelled in the examples in (61) and (61); hence disjunction licenses the conjunctive entailment associated with inclusive-or. There is a second mismatch between logic and language in (61) and (62). The downward entailing expressions in (61) and (62) (negation and the universal quantifier) appear outside the clause that contains disjunction. This is probably another linguistic universal. If negation and disjunction were clausemates, however, then the conjunctive entailment is not enforced in languages like Japanese, Chinese, Hungarian, and many others. In another class of languages, which includes English and German, among others, disjunction yields conjunctive entailments regardless of the placement of the downward entailing operator with respect to the disjunction operator. In these languages, conjunctive entailments of disjunction arise so long as the downward entailing operator c-commands disjunction. This cross-linguistic variation in where downward entailing expressions are able to license conjunctive entailments is a parametric option for natural languages with no counterpart in logic. The existence of such parameters, which partition languages in ways that logic does not even suggest, constitutes one of the strongest arguments for Universal Grammar and against the experience-dependent approach to language development (see Crain, Gualmini and Pietroski 2005 for further discussion).

The next linguistic phenomenon we describe is the interpretation of focus operators, such as only in English. This construction is of particular interest, for several reasons. First, the linguistic phenomenon is (more-than-likely) universal. Second, it has no counterpart in logic. Finally, the interpretation of focus operators involves the generation of a hidden meaning component that has no apparent evidential basis in children’s experience, so this phenomenon represents yet another challenge to the experience-based approach to language development.

7. Hidden Entailments

As proposed by Horn (1969), the meaning of a sentence with the focus operator only, such as (63), can be decomposed into two conjoined propositions. The first proposition pertains to the focus element, Bruce. The content of this proposition is the truthconditional meaning of the original sentence, absent the focus operator, only. So, (64) represents the first meaning component of (60). We will call this its presupposition.

(63) Only Bruce speaks a Romance language.

(64) Bruce speaks a Romance language.

The second meaning component is a proposition that is entailed by (63). The content of this proposition comes from the focus operator only. The entailment is that the property being attributed to the individual in focus (speaking a Romance language) is not a property of anyone else in the conversational context. So, the second meaning component of (63) can be represented as (65). We will call this the assertion.5

(65) For all x [$x \neq$ Bruce], it is not the case that x speaks a Romance language.

Now let us ask if only is downward entailing. As we saw, downward entailing expressions endorse inferences from claims about sets of things to claims about subsets of those things. The entailment from a set to its subsets does not hold for sentences with the focus operator only, however. Consider example (66).

(66) Only Bruce speaks a Romance language.

(67) #Only Bruce speaks French.

English speakers typically deny that (66) entails (67), on the grounds that Bruce might speak Spanish or Italian, and not French. Based on observations such as this, von Fintel (1999) argues that the first meaning component of only, the presupposition, does not contain a downward entailing operator. This explains why the standard diagnostic for downward entailment fails. For example, Bruce speaks a Romance language does not entail Bruce speaks French. What about the second meaning component, the assertion? It turns out that, in this meaning component, sentences with the focus operator only do license inferences from a set to its subsets. For the example in (66) entails (68).

There is general agreement that (65) is entailed by (63), but the status of the meaning component represented in (64) is more controversial: Horn (1969; 1996) and von Fintel (1999) argue that it is a presupposition, whereas Atlas (1993; 1996) and Herburger (2000) take it to be the assertion of the sentence. We will adopt Horn's terminology, and will refer to the first proposition, about the element in focus, as the "presupposition"; and we will refer to the second proposition, about the individuals being contrasted with the element in focus, as the "assertion." We intend no theoretical commitments by adopting this terminology.

(68) For all x [x ≠ Bruce], it is not the case that x speaks French (All the others being contrasted with Bruce do not speak French) So, the second meaning component of the focus operator only apparently does contain a downward entailing expression. We assume that the downward entailing expression is negation, or the semantic equivalent of negation. Thus, although sentences with only (e.g., 66) lack an overt downward entailing expression, they contain a covert downward entailing expression. The covert downward entailing expression appears in the assertion. The acquisition of the covert meaning component of sentences with a focus operator represents another challenge to be confronted by the experience-dependent approach to language development.

Another test of downward entailment, as we saw, is the licensing of conjunctive entailments when disjunction appears in the scope of a downward entailing expression. Let's apply this test to sentences with the focus operator only. Consider (69).

(69) Only Bruce speaks French or Spanish.

(70) #Only Bruce speaks French and only Bruce speaks Spanish.

(71) For all x [x ≠ Bruce], it is not the case that x speaks French and it is not the case that x speaks Spanish

Notice first that (69) does not license the conjunctive entailment in (70). There are circumstances in which (66) is true, but where (70) is false, such as the circumstances in which Bruce speaks just French, or just Spanish, but not both. So, disjunction in the presupposition of sentences with only does not create conjunctive entailments, hence only is not downward entailing in this meaning component. By contrast, (69) entails the conjunctive statement in (71). Hence only passes another test for being downward entailing, but just in one of its meaning components. We might say that the focus operator only is 'partially' downward entailing.

Recent experimental research has sought to determine whether or not children know that sentences with the focus operator only contain a hidden downward entailing operator (negation, or its semantic equivalent). As noted earlier, 4- to 5-year-old children appear to know that or licenses conjunctive entailments in certain downwardentailing contexts, e.g., under negation, and in the Restrictor of the universal quantifier every. So, children's interpretation of or can be used to assess their knowledge of the semantics of only (Goro, Minai & Crain 2004). It seems unlikely that there is relevant evidence in the input about the entailment of sentences with only. On the other hand, if children do not acquire knowledge of the entailment from experience, then children should have to access this knowledge regardless of differences in the language they are learning.

With these objectives in mind, experiments were conducted with Englishspeaking children (using sentences with only … or) and with Japanese-speaking children (using ones with dake… ka …). The research strategy was to investigate their interpretations of disjunction in the overt meaning component and in the covert downward entailing component of sentences with the focus operator only/dake. One of the test sentences is given in (72).

(72) a) Only Bunny Rabbit will eat a carrot or a green pepper. b) Usagichan-dake-ga ninjin ka piiman-wo taberu-yo. rabbit-only-NOM carrot or green pepper-ACC eat-dec Under the decomposition analysis, the meaning of (72) can be partitioned into the two conjoined propositions in (73).

(73) a. Presupposition: Bunny Rabbit will eat a carrot or/ka a green pepper b. Assertion: Everyone other than Bunny Rabbit will not eat a carrot or/ka a green pepper

Within the presupposition component, the disjunction operator or yields disjunctive truth conditions: Bunny Rabbit will eat a carrot or will eat a green pepper. Suppose, first, that children assign the correct interpretation to or within the presupposition component. If so, children should avoid the conjunctive entailment of disjunction in the presupposition, so they should accept (72) in the situation where Bunny Rabbit ate a carrot but not a green pepper. The entire truth conditions are schematically represented in (74).

(74) Condition I Carrot Green Pepper Winnie the Pooh * * Bunny Rabbit √ * Cookie Monster * *

By contrast, within the assertion meaning component of (69), or appears in a downward-entailing environment. Therefore, it licenses the conjunctive entailment -- that everyone else will not eat a carrot and they will not eat a green pepper. Consequently, if children assign the correct interpretation to or within the assertion, they should reject (72) in the situation represented in (75) on the grounds that Cookie Monster ate a green pepper (while, again, Bunny Rabbit ate a carrot but not a green pepper).

(75) Condition II Carrot Green pepper Winnie the Pooh * * Bunny Rabbit √ * Cookie Monster * √

Summarizing, if English/Japanese children assign the inclusive-or interpretation to or/ka, then they should accept the test sentences in Condition I, but they should reject them in Condition II. By contrast, if children assign a different semantics to or/ka, then they could also accept the test sentence in Condition II. We conducted experiments with English-speaking and Japanese-speaking children, to compare their linguistic behaviour. The experiments in English and Japanese were identical in design, with only minimal changes in some of the toy props. The experiment employed the Truth Value Judgment task. There were two experimenters. One of them acted out the stories using the toy props, and the other manipulated the puppet, Kermit the Frog. While the story was being acted out, the puppet watched along with the child subject. In each trial, the story was interrupted - after the introduction of the characters and a description of the situation - so that the puppet could make a prediction about what he thought would happen. Then, the story was resumed, and its final outcome provided the experimental context against which the subject evaluated the target sentence, which had been presented as the puppet's prediction. The puppet repeated his prediction at the end of each story, and then the child subject was asked whether the puppet's prediction had been right or wrong. Twenty-one English-speaking children (mean age 5;0) participated in the experiment, and twenty Japanese-speaking children (mean age 5;4).

The main finding was that both English-speaking children and Japanesespeaking children consistently accepted the test sentences in Condition I, and consistently rejected the test sentences in Condition II. The two groups of children showed no significantly different behaviour in interpreting disjunction within sentences containing a focus operator, only versus dake. Most crucially for our purpose, the high rejection rate in Condition II shows that children assigned conjunctive entailments to disjunction in the assertion component of the test sentences. This, in turn, suggests that they assigned the same semantics to the disjunction operator in each language, despite the differences in input. Children's consistent rejections of the test sentences in Condition 2 provide evidence that they are computing the covert meaning component that is associated with focus operators. As we saw, the covert meaning component expresses a (negative) proposition about a set of individuals that are being contrasted with the element in focus. The findings clearly establish children's ability to compute such contrast sets, although this ability has been questioned by some researchers (cf. Paterson, Liversedge, Rowland and Filik 2003).

In our experiments, there was no evidence of a significant effect of input on the acquisition of disjunction. Both English-speaking children and Japanese-speaking children were able to compute the derived logical truth conditions of disjunction. The experience-dependent account (e.g., Tomasello 2000, 2003) would be hard-pressed to explain the findings of the present studies. In particular, it is hard to see how the experience-dependent account could explain the fact that the same lexical item is interpreted in two different ways in the same sentence. No straightforward learning algorithm would do the trick. It is difficult to see how "cut and paste" operations, like those proposed on the experience-dependent account, could be used to explain the dual interpretations of a single expression of disjunction in sentences with focus operators.

We also leave it as a challenge to such models to account for the absence of any impact of input characteristics on the outcome of acquisition. According to the nativist perspective, children are expected to sometimes follow developmental paths to the adult grammar that would be very surprising from a datadriven perspective. Of course, normal children eventually internalise grammars that are equivalent to those of adults. But a child who has not yet achieved a dialect of English can still be speaking a natural language — albeit one that is (metaphorically) a foreign language, at least somewhat, from an adult perspective. And interestingly children often do exhibit constructions that are not available in the local language — but ones that are available in other adult languages. This is unsurprising if children are free to try out various linguistic options (compatible with Universal Grammar) before 'setting parameters' in ways that specify a particular natural language grammar. This proposal about the course of language development is referred to as the Continuity Hypothesis (Pinker 1984; Crain 1991; Crain and Pietroski 2001). According to one version of the Continuity Hypothesis, child language can differ from the local adult language only in ways that adult languages can differ from each other (Crain 1991; Crain and Pietroski 2001). The idea is that at any given time, children are speaking a possible human language, just not necessarily the particular language that is being spoken around them. Such mismatches between child and adult language are seen to be among the strongest arguments for a Universal Grammar. We first discuss one example of continuity in syntactic development, and then we discuss an example of continuity in semantic development.

In most English whquestions (i.e., questions that begin with wh-words: why, what, where, who, etc.) the wh-question word must be immediately followed by an inflected auxiliary verb (i.e., a tensed form of be, do, can, have, etc.). Hence, the examples in (76a-d) are acceptable, whereas the examples in (76e-h) are not acceptable.

(76) a) Why are you here? b) What do you want to do? c) Where is he going? d) Who don’t you want to win the game? e) *Why you are here? f) *What you want to do? g) *Where he is going? h) *Who you don’t want to win the game?

In Italian, the wh-word corresponding to English why is perché. Italian perché differs from other Italian wh-words in simple questions (for analysis, see Rizzi 1997). As the example in (77) illustrates, the adverb già as well as an entire subject phrase (I tuoi amici) can intervene between perché and the inflected verb (hanno). No linguistic material can intervene with other wh-words in Italian.

(77) Perché (I tuoi amici) già hanno finito il lavoro? Why (the-pl your friends) already have-3pl finished the-sg work 'Why (your friends) already have finished the work?'

However, in complex wh-questions with perché, the intervention of short adverbs or a subject phrase is prohibited in questions like (78) (if the question is asking about the reason for someone's resignation.)

(78) Perché ha detto che si dimetterà? Why have-3sg said that self resign-3sg/future 'Why did he say that he would resign?'

So, complex questions like (78) pattern the same way in both English and in Italian, whereas the simple questions differ, at least for the question words corresponding to 'why.' In both languages, the inflected verb must immediately follow the question word in the complex question. In studies of child English, it has frequently been noted that children produce non-adult why-questions. More specifically, children's simple why questions are often followed by a subject phrase, as in (76e-h). Moreover, such nonadult why-questions persist in children's speech well after they consistently produce adult questions with other wh-words. Adopting the Continuity Hypothesis, Thornton (2004) suggested that children of English-speaking adults initially treat the question-word why in the same way as Italian adults treat perché (cf. de Villiers 1990). If this is correct, an Englishspeaking child should differ from English-speaking adults in the way she forms simple why-questions, but the child should parallel English-speaking adults in producing wellformed long-distance why-questions. From a data-driven perspective, this pattern is surely not anticipated. Since simple questions are more frequent in the input, these should become adult-like in advance of more complex questions, all other things being equal.

Until recently, only data corresponding to children's simple wh-questions were available, so it was difficult to adjudicate between the experience-based account and an account based on the Continuity Assumption. In an experimental and longitudinal diary study, however, Thornton (2004) recorded both simple and complex questions by one child, AL, between the ages of 1;10 and 4;6. By age 3;4, AL produced adult-like whquestions for all wh-words except for why. Nonadult why-questions persisted in (over 80% of) AL's simple questions for more than a year after that, as illustrated in (79).

(79) Why you have your vest on? Why she's the one who can hold it? Why it's his favourite time of day?

What about AL's complex wh-questions with why? From the time AL was 3-years-old until she reached her fourth birthday, she produced 65 complex why-questions, and only seven of them were nonadult. The remaining 58 were adult-like, as were all 39 of AL's complex questions with wh-words other than why. Some examples are provided in (80).

(80) Why do you think you like Cat in the Hat books? Why do you think mummy would not wanna watch the show? What do you think is under your chair? How do you think he can save his wife and her at the same time?

In short, the production data suggest that an English-speaking child analyses whyquestions like the corresponding questions are analyzed in Romance languages, such as perché in Italian. In producing simple why-questions, moreover, AL was ignoring abundant evidence in the input indicating a mismatch between her grammar and that of adult speakers in the same linguistic community. However, AL adhered to the grammatical principles that govern all natural languages, producing adult-like complex why-questions, but nonadult simple why-questions. See Thornton (2004) for several further parallels between AL's why-questions and those of adult speakers of Italian; see Rizzi (1997) for an analysis of questions in Italian.

An example of continuity in semantic development is based on an observation by Goro (2004) who notes that, in Japanese, simple negative sentences with disjunction do not license conjunctive entailments. We noted earlier that Japanese does indeed generate conjunctive entailments for disjunction. It turns out, however, that simple negative sentences lack the conjunctive entailments associated with de Morgan’s laws, at least for adults. In Japanese, for example, the translation of the English sentence Max's computer did not come with Ichat or Isync asserts that Max's computer didn't come with Ichat or it didn't come with Isync; the ‘not both’ reading, rather than the ‘neither’ reading. As a further example, adult Japanese-speakers interpret (81) to mean that the pig didn’t eat the carrot or didn’t eat the pepper. Despite the appearance of ka under negation in surface syntax, ka is interpreted by adults as if it has scope over negation.

(81) Butasan-wa ninjin ka pi’iman-wo tabe-nakat-ta pig-TOP pepper or carrot-ACC eat-NEG-PAST Literally: ‘The pig didn’t eat the pepper or the carrot’ Meaning: ‘The pig didn’t eat the pepper or the pig didn’t eat the carrot’

Based on considerations of language learnability, Goro (2004) hypothesized that Japanese-speaking children would, nevertheless, interpret the disjunction operator ka as licensing conjunctive entailments in simple negative sentences like (81). The Japanese employs sentences with a ~mo ~mo construction, which is semantically similar to the use of conjunction (and) in English. Prediction was that Japanese-speaking children would interpret such sentences in the same way as English-speaking children and adults, despite the absence of this interpretation for adult speakers and, hence, the absence of evidence for this interpretation in the input to children.

In brief, Goro's proposal is that the semantics of natural language disjunction is innately specified as inclusive-or. However, the interaction of disjunction with negation is subject to cross-linguistic variation, as proposed by Szabolcsi (2002). In one class of languages, including English and German, disjunction may be interpreted under local negation, whereas it must be interpreted outside the scope of local negation in another class of languages, including Japanese and Hungarian, regardless of its surface position in such languages. To adopt some technical terminology, Goro proposed that disjunction is a positive polarity item in Japanese (like some in English), but not in English. By definition, a positive polarity item must be interpreted as if it were outside the scope of negation, rather than in its scope. In Japanese, then, the disjunction operator ka appears to have the truth conditions associated with exclusive-or (not both) in simple negative sentences, whereas or creates conjunctive entailments (neither) in the corresponding sentences of English.

Goro's next observation was that the alternative values of the 'positive polarity parameter' for disjunction stand in a subset/superset relation, with English exemplifying the subset value of the parameter, and Japanese exemplifying the superset value. He reasoned that this situation would lead to a 'subset problem' unless children acquiring Japanese initially select the parameter value corresponding to English. Since adult speakers do not make such entailments, it is unlikely that children learn to make them based on the adult input. The reason children should appear more logical than adults in Japanese, Goro suggested, is that children adhere to a principle of language acquisition: the semantic subset principle (Crain, Ni and Conway 1994).

The semantic subset principle enforces an ordering on the values of certain parameters, where one value makes a sentence true in a subset of the circumstances that make it true on the other value. The semantic subset principle compels children to adopt the subset value of the parameter as their initial interpretation; this value is abandoned only on the basis of positive evidence in the local language. If children adopted the superset value instead, they would generate sentences that are not in the local language, in addition to sentences in the local language. This raises a familiar learnability problem: in the absence of negative evidence, it is difficult to see how children would purge their grammars of the means for generating sentences that are not acceptable in the local language. To avoid this problem, the semantic subset principle orders the value of parameters.

To investigate this solution to the 'logical problem of language acquisition,' Goro and Akiba (2004) examined Japanese children's interpretation of negated disjunctions in sentences like (81) using the Truth Value Judgment Task. They interviewed thirty Japanese-speaking children (mean age 5;3) as well as a control group of Japanese-speaking adults. On a typical trial, subjects were asked to judge whether or not (81) was an accurate description of a situation in which the pig had eaten the carrot but not the green pepper. The findings were precisely as anticipated. Japanese-speaking adults uniformly accepted the target sentences (such as 81), whereas children rejected them 75% of the time. The findings are even more compelling once the data from four children, who responded like adults, were set aside. The remaining 26 children rejected the target sentences 87% of the time.

The pattern of responses by Japanese-speaking children are difficult to explain on a "data driven" account of language development, since Japanese-speaking children interpreted negated disjunctions as licensing conjunctive entailments, whereas Japanese-speaking adults did not. On the other hand, the findings are consistent with the continuity hypothesis, according to which child language is expected to diverge from the local adult language, but only in ways that adult languages can differ from each other (see, e.g., Crain 1991, 2002; Crain, Goro and Thornton, in press; Crain and Pietroski 2001, 2002; Thornton 1990, 2004). If children acquired the semantics of the disjunction operator from experience using general-purpose learning algorithms, the fact that ka in Japanese receives a ‘not both’ interpretation in both positive and simple negative sentences would be expected to affect the acquisition process. Specifically, adult input with ka in both positive sentences and in simple negative sentences could mislead Japanese children, prompting them to conclude that ka is a “non-logical” connective, i.e., one that does not obey De Morgan’s laws. Fortunately, this does not happen. The theory of Universal Grammar anticipates that children learning any language should interpret disjunction as inclusive-or, regardless of the input children encounter. The fact that this was found in a language in which the input from adults violates De Morgan’s laws provides further evidence for the continuity assumption (see Jing, Crain & Hsu 2005 for a similar analysis of child Chinese).

CONCLUSION

In this chapter, we have tried to give readers a flavour of some past and current research in the acquisition of syntax and semantics. The presentation of research findings was framed within what is arguably the most central debate in the field: the degree to which human language acquisition is “data driven” or experience-dependent, and the degree to which it is determined by the human genome. Of course, no definitive answer was offered here, one way or the other. At most we have pointed out concepts and consequences of both the nativist approach and the experience-dependent approach to language development. In our view, the experience-dependent approach is implausible in several respects. First, its viability depends on the abilities of children to keep highly detailed records of attested structures in the input. Second, it seems incapable of explaining both language-specific and cross-linguistic generalizations, both in syntactic and in semantic development. Third, it fails to explain how children acquire the 'hidden' meanings of sentences with focus operators, such as only. Fourth, it fails to explain the universal mastery of certain aspects of syntax (c-command) and semantics (downward entailment, inclusive-or). Fifth, it lacks an account of children's nonadult linguistic behaviour, both in syntax and in semantics. Until these challenges are met, the nativist approach appears more convincing. But, this is an empirical matter 34 after all. Only the future will tell if the nativist approach or the experience-dependent approach is closer to the truth.

REFERENCES

Abdul, K.W. (2018, October). *Language Production A Presentation for a Course in Psycholinguistics.* University of Babyblon

Adelman, J. S. (Ed.). (2012). *Visual Word Recognition volume 2: Meaning and context, individuals and development*. Psychology Press.

Adelman, J. S. (Ed.). (2012). *Visual word recognition: models and methods, orthography and phonology* (Vol. 1). Psychology Press.

Alic, M. 2020. *Language Development.* URL; http://www.healthofchildren.com/L/Language-Development.html.

Antes, J.R. (1974). The time course of picture viewing. Journal of Experimental Psychology, 103, 62-70

(K & Castelhano, 2007)

(Carter & M.McCarthy, 1998)

Arnold, J. E., Wasow, T., Asudeh, A., & Alrenga, P. (2004). Avoiding attachment ambiguities: The role of constituent ordering. Journal of Memory and Language, 51, 55–70.

Atlas, Jay. D. 1993. The importance of being "Only": Testing the neo-Gricean versus neo-entailment paradigms. Journal of Semantics 10, 301-318.

Atlas, Jay. D. 1996. 'Only' noun phrases, pseudo-negative generalized quantifiers, negative polarity items, and monotonicity. Journal of Semantics 13, 265-328.

Baker, S. A., Isardi, W. J., Golinkoff, R. M., & Petitto, L. (2005). *The perception of handshapes in American Sign Language*. Memory and cognition, 33, 884–904.

Balota, D. A., Cortese, M. J., Sergent-Marshall, S. D., Spieler, D. H., & Yap, M. J. (2004). Visual word recognition of single-syllable words. *Journal of Experimental Psychology: General*, *133*(2), 283.Spivey, M., Joanisse, M., & McRae, K. (Eds.). (2012). *The Cambridge handbook of psycholinguistics*. Cambridge University Press.

Balota, D. A., Yap, M. J., & Cortese, M. J. (2006). Visual word recognition: The journey from features to meaning (a travel update). In *Handbook of psycholinguistics* (pp. 285-375). Academic Press.

Bock, J. K. (1986a). Meaning, sound, and syntax: Lexical priming in sentence production. Journal of Experimental Psychology: Learning, Memory, and Cognition, 12, 575–586.

Bock, J. K., & Levelt, W. J. M. (1994). Language production: *Grammatical encoding*. In: M. A. Gernsbacher (Ed.), Handbook of Psycholinguistics (pp. 945–984). San Diego: Academic Press.

Boomer, D. S. (1965). Hesitation and grammatical encoding. Language and Speech, 8, 148–158.

Borer, H. and Wexler, K. 1987. The maturation of syntax. In T. Roeper and E. Williams (eds) Parameter Setting. D. Reidel.

Boster, C. and Crain, S. 1994. The interaction of every and or in child language. In Early Cognition and the Transition to Language, University of Texas Working Papers, Austin, TX.

Bowerman, M. F. 1998. The "no negative evidence" problem: How do children avoid constructing overly general grammar? In J. A. Hawkins (ed), Explaining Language Universals. Oxford: Blackwell.

Branigan, H. P., Pickering, M. J., & Cleland, A. A. (2000). Syntactic co-ordination in dialogue. Cognition, 75, B13–B25.

Bresnan, J., & Kaplan, R. (1985). Introduction: Grammars as mental representations of language. In: J. Bresnan (Ed.), The mental representation of grammatical relations (pp. 17–52). Cambridge, MA: MIT Press.

Brown, R., and Hanlon, C. 1970. Derivational Complexity and Order of Acquisition in Child Speech. In J. Hayes (ed), Cognition and the Development of Language. New York: Wiley.

Burke, D. "Language, Aging, and Inhibitory Deficits: Evaluation of a Theory." *Journal of Gerontology: Psychological Sciences* 52B (1997): 254–264

C. D. Elisabeth and P. B. David, July, 27th 2010, Speech perception and production, Carroll, J. B. (1958). Communication theory, linguistics, and psycholinguistics. Review of Educational Research, 28, 79–88.

Chambers, K. E., Onishi, K. H., & Fisher, C. (2003). Infants learn phonotactic regularities from brief auditory experiences. Cognition, 87(2), B69-B77.

Chomsky, N. (1965). Aspects of the theory of syntax. Cambridge, MA: MIT Press.

Corina, D. (1998b). Studies of neural processing in deaf signers: Toward a neurocognitive model of language processing in the deaf. *Journal of Deaf Studies and Deaf Education*, 3(1), 35–48.

Corina, D. P., & Knapp, H. P. (2006). Lexical retrieval in American sign language production. In: L. M. Goldstein, D. H. Whalen, & C. T. Best (Eds), *Papers in laboratory phonology 8:Varieties of phonological competence.* Berlin: Mouton de Gruyter.

Cree, G. S., McNorgan, C., & McRae, K. (2006). Distinctive features hold a privileged status in the computation of word meaning: Implications for theories of semantic memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 32(4), 643–58.

Dr.Chelli.*Languageproduction*.Retrievedfrom*highered.mheducation.com/sites/dl.../9780077122669 _ch12.pdf on Nov 6, 2016.*

Flege, J. E. (1991a). Perception and production: The relevance of phonetic input to L2 phonological learning. In T. Huebner & C. Ferguson (Eds.), Cross-currents in second language acquisition and linguistic theory. Philadelphia: John Benjamins

Frauenfelder, U. H., & Segui, J. (1989). Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory & Cognition*, *17*(2), 134-140.

Frauenfelder, U. H., & Tyler, L. K. E. (1987). Spoken word recognition. The MIT Press: Cambridge, Masshusetts.

Gaskell, G., & Mirković, J. (Eds.). (2016). Speech perception and spoken word recognition. Psychology Press.

Gernsbacher, M. A. (1994). Handbook of psycholinguistic (1st ed): Academic Press.

Gernsbacher, M., A., and Traxler, M., J. (2006). Handbook of Psyhcolinguistic (2nd Edition). London: Elsevier

Graesser, A. C., Millis, K. K., & Zwaan, R. A. (1997). Discourse comprehension. *Annual review of psychology*, *48*(1), 163-189.

Griffin, M.Z. & Ferreira, S.V. (2006, December). Journal of Handbook of Psycholinguistics:*Properties of Spoken Language Production, (21-59). DOI: 10.1016/B978-012369374-7/50003-1*

Groot, A. M. (1985). Word-context effects in word naming and lexical decision. *The Quarterly Journal of Experimental Psychology Section A*, *37*(2), 281-297.

H. M. L. Shannon and N. C. Howard, 2014, Speech perception as an active cognitive process, Published by PMC.

Kintsch, W. (1991). The role of knowledge in discourse comprehension: A construction-integration model. In *Advances in psychology* (Vol. 79, pp. 107-153). North-Holland.

Kroll, J. F., & De Groot, A. M. (Eds.). (2009). Handbook of bilingualism: Psycholinguistic approaches. Oxford University Press.

Ladefoged, P., Maddieson, I. 1996. The sounds of the world's languages. Oxford: Blackwell.

Li, P., & Yip, M. C. (1998). Context effects and the processing of spoken homophones. In *Cognitive processing of the Chinese and the Japanese languages* (pp. 69-89). Springer, Dordrecht.

Lucas, M. (1999). Context effects in lexical access: A meta-analysis. *Memory & Cognition*, *27*(3), 385-398.

M, Paula., E, Maria. (2005). *Language Development and Education Childern with Varying Language Experience*. 1st. Palgrave Macmillan. New York.

Macaya, M., & Perea, M. (2014). Does bold emphasis facilitate the process of visual-word recognition?. *The Spanish journal of psychology*, *17*.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. Cognition, 25(1-2), 71-102.

McRae, K & Jones, M. 2012. The Oxford Handbook of Cognitive Psychology : *Semantic Memory*.

McRae, K., & Jones, M. 2012. The Oxford Handbook of Cognitive Psychology : *Semantic Memory*.

Morris, R. K. (2006). Lexical processing and sentence context effects. In Handbook of psycholinguistics (pp. 377-401). Academic Press.

Moss, J., Schunn, C. D., Schneider, W., McNamara, D. S., & VanLehn, K. (2011). The neural correlates of strategic reading comprehension: Cognitive control and discourse comprehension. *NeuroImage*, *58*(2), 675-686.

Nahri, V., H. (2019). *Keterlambatan Bicara (Speech Delay) Pada Usia Dini*. Surakarta.

Omori, K. (2011). JMAJ: Diagnosis of Voice Disorders (Vol. 54, No. 4).

P. B. David and R. E. Robert, 2005, The Handbook of Speech Perception, UK, Blackwell Publishing Ltd 3.

Pfau,R.,Steinbach, M.,& Woll, B.(Eds.). (2012). Sing lamguage: *An International handbook* (Vol.37). Walter de Gruyter.

Porcello, T. (2004). Speaking of sound: language and the professionalization of sound-recording engineers. Social Studies of Science, 34(5), 733-758.

Rajan, S. (2018). Apraxias. London: University College London

Rastle, K. (2016). Visual word recognition. In *Neurobiology of Language* (pp. 255-264). Academic Press.

S. J. Michael, M. Ken, & J. F. Marc, 2012, The Cambridge Handbook of Psycholinguistics, New York, Cambridge University Press.

Saffron, J., & Sahni S. (2012). Learning the Sounds of Language. In M. Spivey, K. McRae, & M. Joanisse (Eds.), The Cambridge Handbook of Psycholinguistic (Cambridge Handbooks in Psychology, pp 42-58). Cambridge: Cambridge University Press. Doi:10.1017/CB09781139029377.004

Sekar, S. (2019). https://edukasi.kompas.com/read/2019/09/11/18144861/speech-delay-cara-deteksi-dini-dan-pencegahannya?page=all#page2(09/11/19).

Siles, C. P. C. M. 2020. *Milestones-Speech Language Development 0-12 months.* URL; https://www.icommunicatetherapy.com/child-speech-language/child-speech-language-development/child-speech-language-milestones-stages-of-development/speech-language-development-milestones-0-12-months-1-year/.

Simpson, G. B., Peterson, R. R., Casteel, M. A., & Burgess, C. (1989). Lexical and sentence context effects in word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *15*(1), 88.

Spivey, M., J., McRae, M., K., & Joanisse, M., F. (2012). The Cambridge Handbook of Psycholinguistics. New York: Cambridge University Press.

Spivey, M.J., McRae, K., Joanisse, M.F. (2012). *The Cambridge Handbook of Psycholinguistics*. Cambridge University Press. New york. USA.

Steinberg & V. Sciarini. 2006. *An Introduction to Psycholinguistics.* Second edition. *Pearson Logman*. Edinburgh Gate.

Susanto, S. S. R. 2020. *The Hypotheses of Fla and Childern Language Development*. URL; https://jurnal.uns.ac.id/icalc/article/download/16092/12932.

Swinney, D. A. (1979). Lexical access during sentence comprehension:(Re) consideration of context effects. *Journal of verbal learning and verbal behavior*.

T. J. Matthew & G. A. Morton, 2006, HANDBOOK OF PSYCHOLINGUISTICS 2nd EDITION, USA.

Traxler, J.M. & Gernsbacher, A.N. (Eds). (2006). *HANDBOOK OF PSYCHOLINGUISTICS*, 2nd Ed. Amsterdam:Academic Press

Traxler, J.M. & Gernsbacher, A.N. (Eds). (2006). *HANDBOOK OF PSYCHOLINGUISTICS*, 2nd Ed. Amsterdam:Academic Press.

Traxler, J.M. and Gernbascher, A.M. 2006. *Handbook of Psycolinguistics.* 2nd Edition. Elvesier Inc. Oxford, UK.

Traxler, M. J., & Gernsbacher, M. A. (2006). Handbook of Psycholinguistics Second Editon. Oxford, UK.

Traxler, M., & Gernsbacher, M. A. (Eds.). (2011). Handbook of psycholinguistics. Elsevier.

Traxler, M., & Gernsbacher, M. A. (Eds.). (2011). *Handbook of psycholinguistics*. Elsevier

Van Dijk, T. A., & Kintsch, W. (1983). Strategies of discourse comprehension. New York: New York: academic press.

Warren, P. (2013). Introducing psycholinguistics. Cambridge University Press.

Wingfield, A., and Stine-Morrow, E. A. L. "Language and Speech." In *Handbook of Aging and Cognition*, 2d ed. Edited by F. I. M. Craik and T. A. Salthouse. Mahwah, N.J.: Erlbaum Associates, 2000. Pages 359–416

Yap, M. J., & Balota, D. A. (2009). Visual word recognition of multisyllabic words. *Journal of Memory and Language*, *60*(4), 502-529.

Zwaan, R. A., & Rapp, D. N. (2006). Discourse comprehension. In *Handbook of psycholinguistics* (pp. 725-764). Academic Press.

Book :

1. Handbook of Psycholinguistics second edition by Matthew J. Traxler and Morton Ann Gernsbacher.
2. The Cambridge Handbook of Psycholinguistics by Michael J. Spivey, Ken McRae, and Marc F. Joanisse.

Journal :

1. E-Journal of English Language & Literature, AN ANALYSIS OF FIGURATIVE LANGUAGE IN VIEWS (OPINION COLUMN) OF ONLINE PADANG EKSPRES NEWSPAPER, by Nurhaida, and Leni Marlina, English Department, FBS Universitas Negeri Padang

Resources :

1. Psycholinguistic Research on Figurative Language by Ayşe Selin Zorlu, Emine Saçmalı, Firdevs Sümeyye Tok, and Sümeyye Birgül Meylani.

https://www.academia.edu/34935803/Psycholinguistic_Research_on_Figurative_Language

2. Examples of Figurative Language by Your Dictionary.

https://examples.yourdictionary.com/examples-of-figurative language.htm

3. Figurative Language by Psychology Wiki.

https://psychology.wikia.org/wiki/Figurative_language

4. The comprehension of figurative language: Electrophysiological evidence on the processing of irony, by Max Planck Institute for Cognitive and Brain Sciences, 2009.

https://publishup.unipotsdam.de/opus4ubp/frontdoor/deliver/index/docId/3448/file/regel_diss.pdf

Appendices 1. Lesson Plan

RENCANA PEMBELAJARAN SEMESTER (RPS)
MATA KULIAH PSYCHOLIGUISTICS

<table>
<tr><td></td><td colspan="5">UNIVERSITAS KRISTEN INDONESIA
FAKULTAS KEGURUAN DAN ILMU PENDIDIKAN
PROGRAM STUDI PENDIDIKAN BAHASA INGGRIS</td></tr>
<tr><td colspan="6">RENCANA PEMBELAJARAN SEMESTER</td></tr>
<tr><td>MATA KULIAH</td><td>KODE</td><td>RUMPUN MK</td><td>BOBOT (SKS)</td><td>SEMESTER</td><td>TANGGAL PENYUSUNAN</td></tr>
<tr><td>PSYCHOLIGUISTICS</td><td></td><td>Mata Kuliah Prodi</td><td>2</td><td>6</td><td>23 Maret 2020</td></tr>
<tr><td rowspan="2">OTORISASI</td><td colspan="2">Pengembang RPS</td><td colspan="2">Koordinator RMK</td><td>Kaprodi</td></tr>
<tr><td colspan="2">Dr. Lamhot Naibaho, S.Pd., M.Hum., CIQaR</td><td colspan="2">Dr. Lamhot Naibaho, S.Pd., M.Hum., CIQaR</td><td>Hendrikus Male, S.Pd., M.Hum.</td></tr>
</table>

Capaian Pembelajaran (CP)	CPL (CAPAIAN PEMBELAJARAN LULUSAN)	
	 S1 S5 S7 S8 S10 S11	Sikap 1. Bertaqwa kepada Tuhan Yang Maha Esa dan mampu menunjukkan sikap religius. 2. Bekerja sama dan memiliki kepekaan sosial serta kepedulian terhadap masyarakat dan lingkungan. 3. Menginternalisasi nilai, norma, dan etika akademik. 4. Menunjukkan sikap bertanggungjawab atas pekerjaan di bidang keahliannya secara mandiri. 5. Memiliki budi pekerti yang berlandaskan nilai-nilai kristiani: rendah hati, berbagi dan peduli, disiplin, professional dan bertanggung jawab dalam melaksanakan tugas yang dipercayakan.
	 KU1 KU2 KU3 KU4 KU5 KU7 KU8 KU9 KU10	Keterampilan Umum 1. Mampu menerapkan pemikiran logis, kritis, sistematis, dan inovatif dalam konteks pengembangan atau implementasi ilmu pengetahuan dan teknologi yang memperhatikan dan menerapkan nilai humaniora yang sesuai dengan bidang keahliannya. 2. Mampu menunjukkan kinerja mandiri, bermutu, dan terukur. 3. Mampu mengkaji implikasi pengembangan atau implementasi ilmu pengetahuan dan teknologi yang memperhatikan dan menerapkan nilai humaniora sesuai dengan keahliannya berdasarkan kaidah, tata cara dan etika ilmiah dalam rangka menghasilkan solusi, gagasan , desain atau kritik seni, menyususn deskripsi saintifik hasil kajiannya dalam bentuk skripsi atau laporan tugas akhir, dan mengunggahnya dalam laman perguruan tinggi. 4. Menyusun deskripsi saintifik hasil kajian tersebut di atas dalam bentuk skripsi atau laporan tugas akhir, dan mengunggahnya dalam laman perguruan tinggi. 5. Mampu mengambil keputusan secara tepat dalam konteks penyelesaian masalah di bidang keahliannya, berdasarkan hasil analisis informasi dan data. 6. Mampu bertanggung jawab atas pencapaian hasil kerja kelompok dan melakukan supervisi dan evaluasi terhadap penyelesaian pekerjaan yang ditugaskan kepada pekerja yang berada di bawah tanggung jawabnya. 7. Mampu melakukan proses evaluasi diri terhadap kelompok kerja yang berada di bawah tanggung jawabnya, dan mampu mengelola pembelajaran secara mandiri. 8. Mampu mendokumentasikan, memyimpan, mengamankan, dan menemukan kembali data untuk menjamin kesahihan dan mencegah plagiasi.

	KU11 KU12 KK1 P1	9. *Mampu mengungkapkan ide, opini, dan simpulan yang diperoleh dari berbagai sumber ke dalam karya ilmiah untuk menunjang dan mengembangkan kegiatan belajar dan pembelajaran bahasa Inggris.* 10. *Mampu menerapkan teknologi informasi dan komunikasi sebagai media atau sumber pembelajaran, mendukung proses dan pengembangan pembelajaran bahasa Inggris.* 12. *Mampu mengkaji masalah-masalah dalam pembelajaran bahasa Inggris atau implementasi ilmu pendidikan bahasa Inggris berdasarkan kaidah dan etika ilmiah dan menyajikan gagasan atau desain pembelajaran yang lebih baik, dan/atau solusi terhadap masalah dalam pembelajaran bahasa Inggris, dan mampu menyajikan hasil kajian dalam bentuk laporan tertulis atau karya ilmiah (skripsi).* Keterampilan Khusus 1. Mampu merancang dan melaksanakan pembelajaran untuk memampukan peserta didik menguasai kemahiran berbahasa secara komprehensif sesuai dengan kaidah-kaidah gramatikal yang berlaku dalam bahasa Inggris. Pengetahuan 1. Menguasai konsep teoritis yang berkaitan dengan ilmu bahasa, bunyi-bunyi, pembentukan kata, kalimat, makna tekstual dan kontekstual, bahasa ditinjau dari aspek psikologis dan masyarakat secara komprehensif untuk mendukung pembelajaran dan/atau menghasilkan solusi untuk mengatasi masalah-masalah dalam pembelajaran bahasa Inggris.
	CPMK (CAPAIAN PEMBELAJARAN MATA KULIAH)	
	CPMK 1 CPMK 2 CPMK 3 CPMK 4	a. Mahasiswa memahami teori mengenai The basic concepts of psycholinguistics b. Mahasiswa mampu memahami, mengidentifikasi, berargumen dan mengidentifikasi “Language perception/comprehension”. c. Mahasiswa mampu memahami Language production d. Mahasiswa mampu memahami, membedakan, mengeneralisasi dan menentukan dan mempraktekkan Language acquisition”. e. Mahasiswa mampu memahami, membreakdown, mengorganisir dan mengidentifkasi “Language and mind dan Language disorder”.

<table>
<tr><td></td><td>CPMK 5
CPMK 6</td><td colspan="2">f. Mahasiswa mampu “Second language acquisition”</td></tr>
<tr><td>Deskripsi Singkat MK</td><td colspan="3">Mata kuliah ini membahas tentang bagaimana bahasa diperoleh oleh anak, perkembangan bahasa mulai dari anak hingga dewasa, masalah-masalah dalam berbahasa serta hubungan bahasa dangan otak.</td></tr>
<tr><td>Bahan Kajian</td><td colspan="3">1. The basic concepts of psycholinguistics
2. Language perception/comprehension
3. Language production
4. Language acquisition
5. Language and mind dan Language disorder
6. Second language acquisition</td></tr>
<tr><td rowspan="2">Pustaka</td><td colspan="3">Utama:
1. H. Wind Cowles - Psycholinguistics 101-Springer Publishing Company (2010).
2. Katja F. Cantone (auth.) - Code-Switching in Bilingual Children-Springer Netherlands (2007).
3. Susan M. Gass, Larry Selinker - Second Language Acquisition_ An Introductory Course-Routledge (2008).
4. Jean Aitchison - The Articulate Mammal. Intro to Psycholinguistics-Routledge (2007)</td></tr>
<tr><td colspan="3">Penunjang:
1. Online Journal</td></tr>
<tr><td rowspan="2">Media Pembelajaran</td><td colspan="2">Perangkat lunak:</td><td>Perangkat keras:</td></tr>
<tr><td colspan="2">MS Windows
MS Office Power Point
MS Windows Media Player
Internet Google Chrome
Online learning media (teams, zoom, google hangout, and google meet)</td><td>Laptop
Board marker
Whiteboard
Poster
LCD
CD-EIM Student’s Book 3
CD-EIM Student’s workbook 3</td></tr>
</table>

Nama Dosen		Dr. Lamhot Naibaho., S.Pd., M.Hum., CIQaR						
Matakuliah syarat		Linguistics I, II dan III						
Mg Ke-	Sub-CP-MK (Kemampuan Akhir yang Direncanakan)	Bahan Kajian (Materi Pembelajaran)	Bentuk dan Metode Pembelajaran (Media dan sumber belajar)	Estimasi Waktu	Pengalaman Belajar Mahasiswa	Penilaian		
						Kriteria & Bentuk	Indikator	Bobot
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
1-2	Mahasiswa mampu memahami dan mengaplikasikan secara comprehensive Rencana Pembelajaran Semester yang sudah disusun oleh Dosen.	• RPS _ Psycholiguistics • Pembentukan kelompok Belajar	• Bentuk Kuliah Aktifitas di Kelas • Ceramah, Tanya jawab	3 x 50 mnt (hari pertama)	• Terjadi interaksi dan saling berbagi antara dosen dan mahasiswa melalui keterangan dan diskusi serta tanya jawab yang dilakukan oleh dosen dan juga mahasiswa. • Mahasiswa diberikan kebebasan untuk membentuk kelompok belajar sendiri.	• Keterlibatan dalam diskusi • Kualitas dan ketajaman pertanyaan • Kualitas jawaban/respon terhadap pertanyaan.	• Ketepatan dalam memahami informasi yang disampaikan oleh dosen dan forum	5%

	Mahasiswa mampu memahami kajian-kajian yang menjadi cakupan dari mata kuliah Sociolinguistics dan The Role of Psycholiguistics in Language Teaching	• The nature of psycholinguistics.	• Bentuk Kuliah Aktifitas di Kelas • Ceramah, Diskusi kelompok dan Tanya jawab	3 x 50 mnt (hari kedua)	• Terjadi interaksi dan saling berbagi antara dosen dan mahasiswa melalui keterangan dan diskusi serta tanya jawab yang dilakukan oleh dosen dan juga mahasiswa.	• Keterlibatan dalam diskusi • Kualitas dan ketajaman pertanyaan • Kualitas jawaban/respon terhadap pertanyaan.	• Ketepatan dalam memahami informasi yang disampaikan oleh dosen	
3-4	Mahasiswa mampu memahami, berargumen dan mengidentifikasi “Language perception/comprehension”	• “Construction process • Utilization process” (Referensi 1, 2, 3, 4)	• Bentuk Kuliah Aktifitas di Kelas • Presentasi • Diskusi kelompok • Tanya jawab	3 x 50 mnt (hari ke tiga) 3 x 50 mnt (hari ke empat)	• Terjadi interaksi yang baik melalui berbagi pengetahuan tentang “Language and Society dan Language, Dialects, and Varieties” diantara mahasiswa dengan mahasiswa, mahasiswa dan	• Keterlibatan dalam diskusi • Kualitas dan ketajaman pertanyaan • Kualitas jawaban/respon terhadap pertanyaan dalam diskusi • Paper • Slides • Ketepatan dalam menjelaskan	• Ketepatan dalam, memahami, menyampaikan dan menggali informasi yang disampaikan oleh kelompok belajar	10%

					dosen. Mahasiswa belajar bagaimana membuka dan menutup presentasi, belajar mendengar dan menerima pendapat dari orang lain, dan juga belajar bagaimana membuat slide presentasi.			
5-6	Mahasiswa mampu memahami "Language production"	• "Planning what to say Execution the plan" (Referensi 1, 2, 3, 4)	• Bentuk Kuliah Aktifitas di Kelas • Presentasi • Diskusi kelompok • Tanya jawab	3 x 50 mnt (hari ke lima) 3 x 50 mnt (hari ke enam)	• Terjadi interaksi yang baik melalui berbagi pengetahuan tentang "Linguistic Varieties in Multilingual Nations dan *Language Choice in*	• Keterlibatan dalam diskusi • Kualitas dan ketajaman pertanyaan • Kualitas jawaban/respon terhadap pertanyaan dalam diskusi • Paper	• Ketepatan dalam, memahami, menyampaikan dan menggali informasi yang disampaikan oleh kelompok belajar	10%

					Multilingual Communities (including Diglossia, Bilingualism, Multilingualism, Interference, Code-Choice, Code-Switching, and Code-Mixing)" diantara mahasiswa dengan mahasiswa, mahasiswa dan dosen. Mahasiswa belajar bagaimana membuka dan menutup presentasi, belajar mendengar dan menerima pendapat dari orang lain, dan	• Slides • Ketepatan dalam menjelaskan		

					juga belajar bagaimana membuat slide presentasi.			
7-8	Mahasiswa mampu mengerjakan project yang diberikan oleh dosen terkait dengan topik yang sudah dibahasa	**PROJECT 1 Directions: Look at the direction at the Project Sheet**	• Bentuk Kuliah di luar Kelas • Aktifitas Mahasiswa belajar mandiri/kelompok dalam menyelesaikan project yang sudah ditugaskan untuk dikerjakan selama minggu ke 7 dan ke 8	6 x 50 mnt (hari ke 7 dan ke 8)	• Mahasiswa diberikan kesempatan untuk mengalami pembelajaran mandiri/kelompok yang terjadi di luar kelas, serta mengembangkan dan mengekspolarasi ide-ide yang mereka miliki dalam menyelesaikan project 1 yang sudah ditugaskan kepada mereka, serta kemampuan untuk mengelola waktu dan	• Keterlibatan dalam diskusi • Kualitas dan ketajaman hasil Project 1 • Paper • Slides • Ketepatan dalam menjelaskan	• Ketepatan dalam, memahami, menyampaikan dan menggali informasi yang disampaikan oleh kelompok belajar	20%

					aktifitas mereka baik secara individu ataupun kelompok			
9-10	Mahasiswa mampu memahami, membedakan, mengeneralisasi Language acquisition	"Phonological Development • Semantic Development • Syntactic Development" (Referensi 1, 2, 3, 4)	• Bentuk Kuliah Aktifitas di Kelas • Presentasi • Diskusi kelompok • Tanya jawab	3 x 50 mnt (hari ke 9) 3 x 50 mnt (hari ke 10)	• Terjadi interaksi yang baik melalui berbagi pengetahuan tentang "Style, Context, and Register" diantara mahasiswa dengan mahasiswa, mahasiswa dan dosen. Mahasiswa belajar bagaimana membuka dan menutup presentasi, belajar mendengar dan menerima pendapat dari	• Keterlibatan dalam diskusi • Kualitas dan ketajaman pertanyaan • Kualitas jawaban/respon terhadap pertanyaan dalam diskusi • Paper • Slides • Ketepatan dalam menjelaskan	• Ketepatan dalam, memahami, menyampaikan dan menggali informasi yang disampaikan oleh kelompok belajar	10%

					orang lain, dan juga belajar bagaimana membuat slide presentasi.			
11-12	Mahasiswa mampu memahami, membreakdown, mengorganisir dan mengidentifkasi "Language and mind dan Language disorder"	Language process in the brain dan Aphasia (Referensi 1, 2, 3, 4)	• Bentuk Kuliah Aktifitas di Kelas • Presentasi • Diskusi kelompok • Tanya jawab	3 x 50 mnt (hari ke 11) 3 x 50 mnt (hari ke 12)	• Terjadi interaksi yang baik melalui berbagi pengetahuan tentang "Regional and Social Dialects dan Language, Change, Language and Culture" diantara mahasiswa dengan mahasiswa, mahasiswa dan dosen. Mahasiswa belajar bagaimana membuka dan menutup	• Keterlibatan dalam diskusi • Kualitas dan ketajaman pertanyaan • Kualitas jawaban/respon terhadap pertanyaan dalam diskusi • Paper • Slides • Ketepatan dalam menjelaskan	• Ketepatan dalam, memahami, menyampai kan dan menggali informasi yang disampaikan oleh kelompok belajar	10%

					presentasi, belajar mendengar dan menerima pendapat dari orang lain, dan juga belajar bagaimana membuat slide presentasi.			
13-14	Mahasiswa mampu memahami, membedakan, berargumen dan mengkritik "Second language acquisition"	"Children vs adults in second language acquisition, Second language teaching dan Bilingualism" (Referensi 1, 2, 3, 4)	• Bentuk Kuliah Aktifitas di Kelas • Presentasi • Diskusi kelompok • Tanya jawab	3 x 50 mnt (hari ke 13) 3 x 50 mnt (hari ke 14)	• Terjadi interaksi yang baik melalui berbagi pengetahuan tentang ""Language, Change, Language and Culture, Solidarity and Politeness, Language, Attitude and Applications" diantara mahasiswa dengan	• Keterlibatan dalam diskusi • Kualitas dan ketajaman pertanyaan • Kualitas jawaban/respon terhadap pertanyaan dalam diskusi • Paper • Slides • Ketepatan dalam menjelaskan	• Ketepatan dalam, memahami, menyampai kan dan menggali informasi yang disampaikan oleh kelompok belajar	10%

					mahasiswa, mahasiswa dan dosen. Mahasiswa belajar bagaimana membuka dan menutup presentasi, belajar mendengar dan menerima pendapat dari orang lain, dan juga belajar bagaimana membuat slide presentasi.			
15-16	Mahasiswa mampu mengerjakan project yang diberikan oleh dosen terkait dengan topik yang sudah dibahas	**PROJECT 2 Directions: Look at the direction at the Project Sheet**	• Bentuk Kuliah di luar Kelas • Aktifitas Mahasiswa belajar mandiri/kelompok dalam	6 x 50 mnt (hari ke 15 dan ke 16)	• Mahasiswa diberikan kesempatan untuk mengalami pembelajaran mandiri/kelompok yang terjadi di luar kelas, serta	• Keterlibatan dalam diskusi • Kualitas dan ketajaman hasil Project 2 • Paper • Slides • Ketepatan dalam menjelaskan	• Ketepatan dalam, memahami, menyampaikan dan menggali informasi yang disampaikan	25%

			menyelesaikan project yang sudah ditugaskan untuk dikerjakan selama minggu ke 15 dan ke 16		mengembangkan dan mengekspolarasi ide-ide yang mereka miliki dalam menyelesaikan project 2 yang sudah ditugaskan kepada mereka, serta kemampuan untuk mengelola waktu dan aktifitas mereka baik secara individu ataupun kelompok		oleh kelompok belajar	

SISTEM PENILAIAN

I. PERSYARATAN UMUM

A. Kehadiran:

1. Jumlah kuliah tatap muka per semester yang harus dihadiri oleh mahasiswa/i adalah 16 pertemuan.
2. Batas toleransi kehadiran mahasiswa/i 75 % dari total jumlah pertemuan.
3. Kriteria ketidakhadiran mahasiswa/i adalah: S (sakit) ditandai dengan surat keterangan dokter, I (Ijin) ditandai dengan surat ijin resmi, dan A (Alpa), maksimal 4x pertemuan kelas.
4. Mahasiswa aktif dan parsitipatif mengikuti ibadah keluarga besar UKI dan tidak diperkenankan melakukan kegiatan lain selama ibadah berlangsung.
5. Toleransi keterlambatan perkuliahan (dosen + mahasiswa/i) setiap tatap muka adalah 15 menit. Jika setelah 15 menit dosen + mahasiswa/I tidak hadir maka perkuliahan dibatalkan. (kecuali ada persetujuan atau ada masalah tertentu).
6. Perkulihan dapat dilakukan secara online (online learning*) dengan pemberitahuan info minimal sehari sebelumnya melalui grup WA atau info resmi

B. Perkualiahan:

1. Perkuliahan diawali dan diakhiri dengan doa dan renungan singkat diambil dari renungan singkat harian “Renungan Harian Pagi, Siang dan Malam” secara online
2. Mata kuliah yang dilaksanakan mahasiswa berbasis KKNI.
3. Mata kuliah berbasis KKNI dinilai/dievaluasi per topik yang telah tuntas
4. Persentase penilaian/evaluasi ditentukan oleh dosen yang bersangkutan sesuai kompetensi MK dan capaian pembelajaran.
5. Tidak diperkenankan meninggalkan kelas selama perkuliahan tanpa ijin oleh dosen.

6. Mahasiswa tidak diijinkan membuka HP saat proses belajar mengajar berlangsung tanpa ijin oleh dosen.
7. Mahasiswa memakai busana yang sopan.
8. Tidak membuat kegaduhan selama proses pembelajaran berlangsung.
9. Mata Kuliah ini Mata Kuliah ini dibagi dalam 2 hari dalam seminggu

C. Kejahatan akademik: plagiarisme Menurut Peraturan Menteri Pendidikan RI Nomor 17 Tahun 2010:

"Plagiat adalah perbuatan sengaja atau tidak sengaja dalam memperoleh atau mencoba memperoleh kredit atau nilai untuk suatu karya ilmiah, dengan mengutip sebagian atau seluruh karya dan atau karya ilmiah pihak lain yang diakui sebagai karya ilmiahnya, tanpa menyatakan sumber secara tepat dan memadai." (Permendik No 17 Tahun 2010 dan Panduan Anti Plagiasime terlampir).

Sanksi sesuai Permendik No 17 Tahun 2010 Pasal 12:

1. teguran;
2. peringatan tertulis;
3. penundaan pemberian sebagian hak mahasiswa;
4. pembatalan nilai satu atau beberapa mata kuliah yang diperoleh mahasiswa;
5. pemberhentian dengan hormat dari status sebagai mahasiswa;
6. pemberhentian tidak dengan hormat dari status sebagai mahasiswa; atau
7. pembatalan ijazah apabila mahasiswa telah lulus dari suatu program.

II.PERSYARATAN KHUSUS

A. Tugas dan Tanggung jawab mahasiswa/i

Pada setiap tatap muka mahasiswa/i diwajibkan berpartisipasi aktif dalam proses perkuliahan melalui hal-hal berikut

1. Kuis reguler: mahasiswa wajib menyelesaikan semua tugas yang diberikan oleh dosen.
2. Presentasi: mahasiswa/i wajib berpartisipasi aktif dalam diskusi yang diadakan dalam setiap tatap muka sesuai kebutuhan materi perkuliahan (lihat RPS).
3. Mahasiswa/I wajib berpartisipasi aktif dalam studi mandiri yang dilakukan di luar kampus sesuai topik materi perkuliahaan yang sudah ditentukan dalam RPS.
4. Tugas terstruktur: mahasiswa/i wajib membentuk kelompok untuk mendiskusikan berbagai topik-topik yang sudah disampaikan kepada masing-masing kelompok dengan baik dan menyerahkan hasil diskusi pada setiap pertemuan.

B. Gaya Selingkung Pengerjaan Tugas

1. Untuk mengerjakan tugas Project, mahasiswa/i wajib mematuhi ketentuan berikut:
   a. Artikel mahasiswa/i harus ditulis dengan komposisi: Pendahuluan (1 hal), Pembahasan (Sesuai dengan kejelasan dan ketajaman dari materi yang didiskusikan), Kesimpulan (½ hal).
   b. Daftar referensi minimal menggunakan 3 buku dan 5 jurnal ilmiah.
   c. Pengutipan dan penulisan daftar pustaka menggunakan “APA Style 6th edition” (terlampir).
   d. Ketentuan kertas A4, huruf Times New Roman, ukuran jenis 12, spasi 1½.
2. Untuk mengerjakan tugas makalah kelompok, mahasiswa/i wajib mematuhi ketentuan berikut:
   a. Artikel mahasiswa/i harus ditulis dengan komposisi: Pendahuluan berisi permasalahan dan pentingnya isu/fenomena tersebut dibahas (2 hal), Tinjauan Teoritis berisi teori apa yang hendak digunakan sebagai pisau analisis (Sesuai dengan kejelasan dan ketajaman dari materi yang didiskusikan), Pembahasan (Sesuai dengan kejelasan dan ketajaman dari materi yang didiskusikan), Kesimpulan (1 hal).
   b. Daftar referensi minimal menggunakan 5 buku dan 10 jurnal ilmiah.
   c. Pengutipan dan penulisan daftar pustaka menggunakan “APA 6th edition (American Psychological Association).

d. Ketentuan kertas A4, jenis huruf Times New Roman, ukuran 12, spasi 1½.

III. PENILAIAN

A. *rubric analitik untuk penilaian presentasi

Aspek/ dimensi yg dinilai	Kreteria Penilaian				
	Sangat Kurang (Skor < 20)	Kurang (21-40)	Cukup (41-60)	Baik (61-80)	Sangat Baik (Skor ≥ 81)
Organisasi	Tidak ada organisasi yang jelas. Fakta tidak digunakan untuk mendukung pernyataan.	Cukup fokus, namun bukti kurang mencukupi untuk digunakan dalam menarik kesimpulan	Presentasi mempunyai fokus dan menyajikan beberapa bukti yang mendukung kesimpulankesimpulan.	terorganisasi dengan baik dan menyajikan fakta yang meyakinkan untuk mendukung kesimpulankesimpulan	terorganisasi dengan menyajikan fakta yang didukung oleh contoh yang telah dianalisis sesuai konsep
Isi	Isinya tidak akurat atau terlalu umum. Pendengar tidak belajar apapun atau kadang menyesatkan	Isinya kurang akurat, karena tidak ada data faktual, tidak menambah pemahaman pendengar	Isi secara umum akurat, tetapi tidak lengkap. Para pendengar bisa mempelajari beberapa fakta yang tersirat, tetapi mereka tidak menambah wawasan baru tentang topik tersebut.	Isi akurat dan lengkap. Para pendengar menambah wawasan baru tentang topik tersebut.	Isi mampu menggugah pendengar untuk mengambangka n pikiran.
Gaya Presentasi	Pembicara cemas dan tidak nyaman, dan membaca berbagai catatan daripada berbicara. Pendengar sering diabaikan. Tidak terjadi kontak mata karena pembicara lebih	Berpatokan pada catatan, tidak ada ide yang dikembangka n di luar catatan, suara monoton	Secara umum pembicara tenang, tetapi dengan nada yang datar dan cukup sering bergantung pada catatan. Kadangkadang kontak mata dengan pendengar diabaikan.	Pembicara tenang dan menggunakan intonasi yang tepat, berbicara tanpa bergantung pada catatan, dan berinteraksi secara intensif dengan pendengar. Pembicara selalu kontak mata dengan pendengar.	Berbicara dengan semangat, menularkan semangat dan antusiasme pada pendengar

	banyak melihat ke papan tulis atau layar.				

B. *Rubrik Holistik

GRADE	SKOR	KRITERIA PENILAIAN
Sangat kurang	<20	Rancangan yang disajikan tidak teratur dan tidak menyelesaikan permasalahan
Kurang	21–40	Rancangan yang disajikan teratur namun kurang menyelesaikan permasalahan
Cukup	41– 60	Rancangan yang disajikan tersistematis, menyelesaikan masalah, namun kurang dapat diimplementasikan
Baik	61- 80	Rancangan yang disajikan sistematis, menyelesaikan masalah, dapat diimplementasikan, kurang inovatif
Sangat Baik	>81	Rancangan yang disajikan sistematis, menyelesaikan masalah, dapat diimplementasikan dan inovatif

C. *Rubrik Skala Persepsi untuk Penilaian Presentasi Lisan

Aspek/dimensi yang dinilai	Sangat Kurang	Kurang	Cukup	Baik	Sangat Baik
	<20	(21-40)	(41-60)	(61-80)	≥80

Kemampuan Komunikasi					
Penguasaan Materi					
Kemampuan menghadapi Pertanyaan					
Penggunaan alat peraga presentasi					
Ketepatan menyelesaikan masalah					

*) sumber rubrik diambil dari *Panduan penyusunan KPT di Era Industri 4.0 tahun 2019*.

IV. Skala nilai akhir dalam huruf dan angka:

Nilai Akhir (NA)	Nilai Huruf (NH)	Nilai Mutu (NM)
80,0-100,0	A	4,0
75,0-79,0	A-	3,7
70,0-74,9	B+	3,3
65,0-69,9	B	3,0
60,0-64,9	B-	2,7
55,0-59,9	C	2,3
50,0-54,9	C-	2,0
45,0-49,9	D	1,0
<44,9	E	0

V. Prosentase Tahap Penilaian Tugas dan kewajiban mahasiswa (dapat diganti/disesuaikan oleh dosen)

Nama Mahasiswa	**Minggu ke**	**TOTAL SCORE**

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	
		5%	**10%**		**10%**		**20%**		**10%**		**10%**		**10%**		**25%**		
1																	
2																	
3																	
4																	
5																	
N																	

Jakarta, 15 Juli 2020

Mengetahui,
Ketua Program Studi,

ttd

Hendrikus Male, M.Hum.

Disusun Oleh
Dosen Pengampu,

Dr Lamhot Naibaho, S.Pd., M.Hum., CIQaR.

Appendices 2: Compiler's Resume Summary

Latest Position	:	Lecturer at English Education Department Faculty of Teacher Training and Education
Lastest Job Function	:	Education
Latest Industry Sector	:	Education

Personal Identity

Name	:	Lamhot Naibaho, S.Pd.,M.Hum
NIDN	:	0118118504/121927
Place and date of Born	:	Buluduri, 18 November 1985
Age	:	35 years old
Gender	:	Male
Functional title	:	III/C-Penata
Academic title	:	Associate Professor
Affiliation	:	Christian University of Indonesia
Office Address	:	Jl. Mayjen Sutoyo No. 2 Cawang, Jakarta Timur
Mobile	:	0812-1225-2045
Home Address	:	*Perumahan Bojong Menteng Blok A. No. 282. Jl. Jati Timur VI Rawalumbu, Bekasi Timur Jawa Barat*
e-mail	:	lamhot.naibaho@uki.ac.id / lnaibaho68@gmail.com

Educational Background

Bachelor Degree	:	English Education Department (State University of Medan) - 2009
Magister	:	English Applied Linguistics Department (State University of Medan) - 2012
Doctoral/Ph.D	:	Language Education Department (State University of Jakarta) - 2016

Research Experiences

1. Improving Students' Narrative Writing Ability through Self-Regulated Strategy Development – University Research
2. The Analysis of English Test Designed by Junior High School Teachers' Using Blooms' Taxonomy – University Research

3. An Analysis of English National Final Exam (UAN) for Senior High School Viewed from Bloom's Taxonomy Theory – University Research
4. The Description of Students' Interest and Learning Achievement on Christian Leadership at *Universitas Kristen Indonesia* – University Research
5. Improving Students' Essay Writing Ability through Consultancy Pre-writing Protocol – University Research
6. The Active Role of Families in Building Students' Character at *Universitas Kristen Indonesia* – University Research
7. Improving Students' Speaking Ability through Independent Learning Method at Christian University of Indonesia – University Research
8. Language Acquisition by A Child Suffering of Language Delay – ***RESEARCH GRANTS from Ministry of Research, Technology and Higher Education of Indonesia***
9. The Evaluation of SCL and Students' Internship Program at *Sekolah Mitra PSKD Se Jakarta* – University Research
10. The Retention and Preservation of Regional Languages as Multi-Cultural Identities of Indonesia in the Globalization Era – University Research
11. The Description of medical students' interest and achievement on anatomy at faculty of medicine *Universitas Kristen Indonesia* – University Research
12. Building Employees' Mental Health: The Correlation between Transactional Leadership and Training Program with Employees' Work Motivation at XWJ Factory – University Research
13. Healthy Work Culture Stimulate Performance - – University Research
14. Analysis of Nursing Quality Services – University Research
15. The Asmat Tribe Perception of Child Parenting – ***GRANTS from Wahana Visi Indonesia***
16. Organizational Development Mentoring and Procurement Skill Training – ***GRANTS from United State of America Ambassy***
17. The Effectiveness of Mastery Learning Technique On Improving Students' Ability in Completing English National Examination – University Research

18. Analysis of Student Morality according to Kohlberg and Lickona's Theory at *Sekolah SMP Negeri 9 dan 29 Bekasi* - ***RESEARCH GRANTS from Ministry of Research, Technology and Higher Education of Indonesia.***
19. *Maintenance and Preservation of Regional Languages as the Multi-Cultural Identity of the Indonesian Nation in Industry 4.0* - ***RESEARCH GRANTS from Ministry of Research, Technology and Higher Education of Indonesia.***
20. *Evaluation of the Scavengers Development Program Conducted by GMIM Getsemani Sumompo* - ***RESEARCH GRANTS from Ministry of Research, Technology and Higher Education of Indonesia***

Scientific Publications

1. *Peranan Guru dalam Pengajaran Bahasa Inggris* JDP (Jurnal Dinamika Pendidikan).
2. *Bersama Mendukung Otonomi Daereh sebagai Langkah Menuju Daerah yang Maju, Masyarakat yang Makmur, Sejahtera dan Sentosa (*APKASI **(Asosiasi Pemimpin Kepala Daerah se Indonesia)/Lomba Penulisan Karya Ilmiah Tingkat S2, S3, Dosen dan Profesor*).***
3. *Peran Pendidikan dan Kebudayaan dalam Pembangunan Karakter dan Peradaban Bangsa Indonesia yang Majemuk (*The Ary Suta Center/Strategic Management).
   Naibaho, L. (2014). Peran Pendidikan dan Kebudayaan dalam Pembangunan Karakter dan Peradaban Bangsa Indonesia yang Majemuk. ***Jurnal the Ary Suta Center Series on Strategic Management*, *27*(0), 69**.
4. Phonological Acquisition of A Child Suffering from Language Delay **(International Journal of Language Education and Culture Review).**
   Naibaho, L. (2016). Phonological Acquisition of A Child Suffering from Language Delay. *International Journal of Language Education and Culture Review*, *2*(1), 33-42
5. Improving Students' Essay Writing Ability through Consultancy Prewriting Protocol at Christian University of Indonesia **(The Asian ESP Journal, 28 Agustus 2018) - *SCOPUS INDEXED JOURNAL.***

Naibaho, L. (2016). Improving Students' Essay Writing Ability through Consultancy Prewriting Protocol at Christian University of Indonesia. *The Asian EFL Journal, 3*, 147-160

6. Teachers'roles On English Language Teaching: A Students Centered Learning Approach **(International Journal of Research-Granthaalayah, 7/4/2019 Page. 206-212.**

Naibaho, L. (2019). Teachers'roles on English Language Teaching: A Students Centered Learning Approach. *International Journal of Research-Granthaalayah*, *7*(4), 206-212

7. Optimizing the Air Transport Operations of Indonesian National Army-Air Force on Overcoming the Impact of the Future Natural Disasters **(Journal of Advances in Social Science and Humanities, 4/2/2019 Page.1-12.**

Suryaningsih, L., Mastra, I. G., & Naibaho, L. (2018). Optimizing the Air Transport Operations of Indonesian National Army-Air Force on Overcoming the Impact of the Future Natural Disasters. *Journal of Advances in Social Science and Humanities*, *4*(2), 1-12

8. The Effectiveness Of Scaffolding Method On Students'speaking Achievement **(International Journal of Research-Granthaalayah, 7/5/2019, Page. 193-201)**

Naibaho, L. (2019). The Effectiveness Of Scaffolding Method On Students'speaking AchievemeNT. *International Journal of Research-Granthaalayah*, *7*(5), 193-201

9. Improving Eight Graders' Reading Comprehension Using Student Team Achievement Division (STAD) at SMP Strada Santo Fransiskus **(EFL Theory & Practice: Voice of EED UKI Page: 90 - 101 (Prosiding)**

Naibaho, L., & Sangga, R. E. (2019). Improving Eight Graders' Reading Comprehension Using Student Team Achievement Division (STAD) at SMP Strada Santo Fransiskus

10. Students' Perception on Guessing Game Use in Learning Vocabulary at SMPK Ignatius Slamet Riyadi **(EFL Theory & Practice: Voice of EED UKI Page 160 - 171 (Prosiding).**

Naibaho, L., & Ambrosia, Y. (2019). Students' Perception on Guessing Game Use in Learning Vocabulary at SMPK Ignatius Slamet Riyadi

11. Implementation Of Students Centered Learning At Persekutuan Sekolah Kristen Djakarta **(International Journal of Engineering Sciences & Research Technology, 7/8/2018 Page. 585-592).**
    Tyas, E. H., & Naibaho, L. (2018). Implementation Of Students Centered Learning At Persekutuan Sekolah Kristen Djakarta. *International Journal Of Engineering Sciences & Research Technology*, *7*(8), 585-592.
12. **12.** Penerapan Sistem Emulsi Membran Terhadap Kontrol Air Tambak Dan Sistem Ijuk Terhadap Penyediaan Air Bersih Di Bumi Dipasena **(Journal Comunità Servizio, 1/1/2019 Page. 19 – 27).**
    Tyas, E. H., & Naibaho, L. (2018). Implementation of Students Centered Learning at Persekutuan Sekolah Kristen Djakarta. *International Journal of Engineering Sciences & Research Technology*, *7*(8), 585-592.
13. The Integration of Group Discussion Method Using Audio Visual Learning Media Toward Students' Learning Achievement On Listening **(International Journal of Research-Granthaalayah, 7/8/2019 Page 438 -445).**
    Naibaho, L. (2019). The Integration of Group Discussion Method Using Audio Visual Learning Media Toward Students'learning Achievement on Listening. *International Journal of Research-Granthaalayah*, *7*(8), 438-445.
14. Use Of Construction Inquiri Learning Model To Improve The Interest Of Learning Students Grade XI SMA Angkasa 2 In Coloid Materials **(PEOPLE: International Journal of Social Sciences, 5/2/2019 Page.908 – 917)**
    Sormin, E., Julianti, K., Nadeak, B., & Naibaho, L. (2019). Use of construction inquiri learning model to improve the interest of learning students grade XI SMA Angkasa 2 in coloid materials. *PEOPLE International Journal of Social Sciences*, *5*(2), 908-917.
15. The Description of Medical Students' Interest and Achievement on Anatomy at Faculty of Medicine Universitas Kristen Indonesia **(International Journal of Sciences: Basic and Applied Research (IJSBAR) 39/1/2018 Page.121-133).**
    Nadeak, B., & Naibaho, L. (2018). The Description of medical students' interest and achievement on anatomy at faculty of medicine

Universitas Kristen Indonesia. *International Journal of Sciences: Basic and Applied Research (IJSBAR)*, *39*(1), 121-133.

16. Building Employees' Mental Health: The Correlation between Transactional Leadership and Training Program with Employees' Work Motivation at XWJ Factory (Indian Journal of Public Health Research & Development, 10/6/2019 Page.1373 -1379) - ***SCOPUS INDEXED JOURNAL.***

    Nadeak, B., Iriani, U. E., Naibaho, L., Sormin, E., & Juwita, C. P. (2019). Building Employees' Mental Health: The Correlation between Transactional Leadership and Training Program with Employees' Work Motivation at XWJ Factory. *Indian Journal of Public Health Research & Development*, *10*(6), 1373-1379.

17. Healthy Work Culture Stimulate Performance (Indian Journal of Public Health Research & Development, 10/6/2019, Page.1379-1386) - ***SCOPUS INDEXED JOURNAL.***

    Nadeak, B., Naibaho, L., Sormin, E., & Juwita, C. P. (2019). Healthy Work Culture Stimulate Performance. *Indian Journal of Public Health Research & Development*, *10*(6), 1385-1389.

18. Analysis of Nursing Quality Services (Indian Journal of Public Health Research & Development, 10/6/2019 Page. 1386 – 1393) - ***SCOPUS INDEXED JOURNAL***.

    Nadeak, B., Simanjuntak, D. R., Naibaho, L., Sormin, E., Juwita, C. P., & Pardede, S. O. (2019). Analysis of Nursing Quality Services. *Indian Journal of Public Health Research & Development*, *10*(6), 1380-1384.

19. The Effectiveness 0f Number Head Together Strategy On Improving Students' English Achievement At Xyz School (International Journal of Research-Granthaalayah, 7/10/2019 Page.362 – 370).

    Naibaho, L. (2019). The Effectiveness of Number Head Together Strategy on Improving Students'english Achievement at XYZ School. *International Journal of Research-GRANTHAALAYAH*, *7*(10), 362-370.

20. The Effectiveness of Independent Learning Method on Students' Speaking Achievement at Christian University of Indonesia Jakarta (Asian EFL Journal Research Articles, 23/6.3/ 2019, Page. 142-154) - ***SCOPUS INDEXED JOURNAL***.

Naibaho, L. (2019). The Effectiveness of Independent Learning Method on Students' Speaking Achievement at Christian University of Indonesia Jakarta. *Asian EFL Journal*, *23*(6), 142-154.

21. Investigating the effect of learning multimedia and thinking style preference on learning achievement on anatomy at *Universitas Kristen Indonesia* (Journal of Physics: Conference Series, 3 Des 2019 1387 (1), 012116) - ***SCOPUS INDEXED JOURNAL.***

Nadeak, B., & Naibaho, L. (2019, November). Investigating the effect of learning multimedia and thinking style preference on learning achievement on anatomy at Universitas Kristen Indonesia. In *Journal of Physics: Conference Series* (Vol. 1387, No. 1, p. 012116). IOP Publishing.

22. The urgency of entrepreneurship learning in the industrial age of 4.0 (Journal of Physics: Conference Series, 3 Des 2019 1387 (1), 012032) - ***SCOPUS INDEXED JOURNAL.***

Tyas, E. H., & Naibaho, L. (2019, November). The urgency of entrepreneurship learning in the industrial age of 4.0. In *Journal of Physics: Conference Series* (Vol. 1387, No. 1, p. 012032). IOP Publishing.

23. Building Superior Human Resources through Character Education (Test Engineering and Management Volume 83 Page 11864 – 11873 Issue March – April 2020) – ***SCOPUS INDEXED JOURNAL***.

Tyas, E. H., & Naibaho, L. (2020). Building Superior Human Resources through Character Education. *TEST Engineering & Management*, *83*, 11864-11873.

24. Enhancing Community Legal Awareness of Land Right Disputes through the Use of Legal Aid Institutions (International Journal of Psychosocial Rehabilitation, Vol. 24, Issue 08, 2020) - ***SCOPUS INDEXED JOURNAL***.

Tehupeiory, A., & Naibaho, L. (2020). Enhancing community legal awareness of land right disputes through the use of legal aid institutions. *International Journal of Psychosocial Rehabilitation*, *24*(8), 1223-1231.

25. The Effectiveness of Problem-Based Learning on Students' Critical Thinking (Jurnal Dinamika Pendidikan, Vol 13, No. 1 2020).

Nadeak, B., & Naibaho, L. (2020). The Effectiveness of Problem-Based Learning on Students'critical Thinking. *Jurnal Dinamika Pendidikan*, *13*(1), 1-7.

26. **The Use of Google on Completing English Assignment by the Students' of English Education Department at Universitas Negeri Medan (International Journal of Research-Granthaalayah, 8/6/2020 Page. 150-155).**

    Digeyasa, I. W., & Naibaho, L. (2020). The Use of Google on Completing English Assignment by the Students'of English Education Department at Universitas Negeri Medan. *International Journal of Research*, *8*(6), 150-155.

27. COVID-19 and Students' Anxiety Management (International Journal of Innovation, Creativity and Change, Volume 13, Issue 7, 2020) - ***SCOPUS INDEXED JOURNAL.***

    Nadeak, B., Naibaho, L., & Silalahi, M. (2020). COVID-19 and Students' Anxiety Management. *International Journal of Innovation, Creativity and Change*, *13*(7), 1574-1587.

28. Building a Culture of Tolerance since Early Childhood **(International Journal of Research-GRANTHAALAYAH 8 (8), 244-249).**

    Tyas, E. H., & Naibaho, L. (2020). Building a Culture of Tolerance since Early Childhood. *International Journal of Research-GRANTHAALAYAH*, *8*(8), 244-249

29. Female and Males' Brain Tendencies In Learning English as A Second Language **(International Journal of Research-Granthaalayah, 8/7/2020 Page. 211-216).**

    Naibaho, L. (2020). Female and Males'brain Tendencies In Learning English as A Second Language. *International Journal of Research-GRANTHAALAYAH*, *8*(7), 211-216.

30. The Effectiveness Of Mastery Learning Technique On Improving Students'ability In Completing English National Examination **(International Journal of Research-GRANTHAALAYAH Vol 8 Nomor 2 2020).**

    Nadeak, B., & Naibaho, L. (2020). The Effectiveness of Mastery Learning Technique on Improving Students'ability In Completing English National Examination. *International Journal of Research-GRANTHAALAYAH*, *8*(2), 57-62.

31. A harmony among of religious community is required amidst the covid-19 pandemic **(International Journal of Research-GRANTHAALAYAH, Vo; 8 No. 9 2020).**

Tyas, E. H., & Naibaho, L. (2020). A harmony among of religious community is required amidst the covid-19 pandemic. *International Journal of Research-GRANTHAALAYAH*, *8*(9), 422-428.

32. Korelasi Kemampuan Berpikir Kritis Mahasiswa dan Penggunaan Media Sosial Terhadap Capaian Pembelajaran Pada Masa Pandemi Covid-19 (Jurnal Konseling dan Pendikan Vol 8, No 2 (2020). **National Accredite Journal _ *SINTA 2.***

Nadeak, B., Juwita, C. P., Sormin, E., & Naibaho, L. (2020). Hubungan kemampuan berpikir kritis mahasiswa dengan penggunaan media sosial terhadap capaian pembelajaran pada masa pandemi Covid-19. *Jurnal Konseling dan Pendidikan*, *8*(2), 98-104

33. Video-Based Learning on Improving Students'learning Output (PalArch's Journal of Archaeology of Egypt/Egyptology, Vol 17. Issue 2. 2020) - ***SCOPUS INDEXED JOURNAL.***

Nadeak, B., & Naibaho, L. (2020). Video-Based Learning on Improving Students'learning Output. *PalArch's Journal of Archaeology of Egypt/Egyptology*, *17*(2), 44-54.

34. HOTS Learning Model Improves the Quality of Education **(International Journal of Research-GRANTHAALAYAH 9 (1), 176-182).**

Tyas, E. H., & Naibaho, L. (2021). HOTS Learning Model Improves the Quality of Education. *International Journal of Research-GRANTHAALAYAH*, *9*(1), 176-182.

35. Motivation and HRM Factors Relation to the Employee Loyalty (Polish Journal of Management Studies Vol. 22 No. 2. 2020) - ***SCOPUS INDEXED JOURNAL.***

Nadeak, B., & Naibaho, L. (2020). Motivation and HRM Factors Relation to the Employee Loyalty. *Polish Journal of Management Studies*, *22*(2).

36. Learning Management in Suburban Schools During the Midst of Covid-19 (Psychology and Education Journal, Vol. 58 n0. 2 2021) - ***SCOPUS INDEXED JOURNAL.***

Nadeak, B., Naibaho, L., Sunarto, S., Tyas, E. H., & Sormin, E. (2021). Learning Management in Suburban Schools During the Midst of COVID-19. *Psychology and Education Journal*, *58*(2), 1131-1139.

37. The Integration of Lecturers' Professionalism and Intelligence with Environment Insight (Psychology and Education Journal, Vol. 58 n0. 2 2021) - ***SCOPUS INDEXED JOURNAL.***

Sunarto, E., Naibaho, L., Sormin, E., & Nadeak, B. (2021). The Integration of Lecturers' Professionalism and Intelligence with Environment Insight. *Psychology and Education Journal*, *58*(2), 5981-5987

38. Lecturers' Cultural Sensitivity on Using Social Media-Facebook and Instagram (Psychology and Education Journal, Vol. 58 n0. 2 2021) - ***SCOPUS INDEXED JOURNAL.***

Tyas, E. H., Sunarto, L. N., Nadeak, B., & Sormin, E. (2021). Lecturers' Cultural Sensitivity on Using Social Media-Facebook and Instagram. *Psychology and Education Journal*, *58*(2), 5974-5980.

39. Strengthening the Values of Christian Education in Facing the New Normal Era. (Psychology and Education Journal, Vol. 58 n0. 2 2021) - ***SCOPUS INDEXED JOURNAL.***

Kolibu, D. R., Peter, R., Naibaho, L., Paparang, S. R., & Hanock, E. E. (2021). Strengthening the Values of Christian Education in Facing the New Normal Era. Psychology and Education Journal, 58(2), 10937-10944.

40. The Integration of Lecturers' Professionalism and Intelligence with Environment Insight. (Psychology and Education Journal, Vol. 58 n0. 2 2021) - ***SCOPUS INDEXED JOURNAL.***

Sunarto, E., Naibaho, L., Sormin, E., & Nadeak, B. (2021). The Integration of Lecturers' Professionalism and Intelligence with Environment Insight. *Psychology and Education Journal*, *58*(2), 5981-5987.

41. Learning Management in Suburban Schools During the Midst of COVID-19 (Psychology and Education Journal, Vol. 58 n0. 2 2021) - ***SCOPUS INDEXED JOURNAL.***

Nadeak, B., Naibaho, L., Sunarto, S., Tyas, E. H., & Sormin, E. (2021). Learning Management in Suburban Schools During the Midst of COVID-19. *Psychology and Education Journal*, *58*(2), 1131-1139.

42. Pre-Service Teachers’ Soft Skills and Achievement (**Turkish Journal of Computer and Mathematics Education (TURCOMAT), Vol. 12 No. 10) - *SCOPUS INDEXED JOURNAL.***

    Naibaho, L. (2021). Pre-Service Teachers’ Soft Skills and Achievement. *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, *12*(10), 491-496.

43. Managing Tri Pusat Pendidikan in the Covid -19 Pan-Demic (**International Journal of Research -GRANTHAALAYAH, Vol.9 Issue 4).**

    Tyas, E. H., & Naibaho, L. (2021). Managing Tri Pusat Pendidikan in the COVID-19 Pandemic. *International Journal of Research-GRANTHAALAYAH*, *9*(4), 492-500.

44. The Analysis of Code-switching Integration Realization on Students Classroom Performance **(Atlantis-Press _ Proceeding published by Springer and indexed by WoS, 2021.**

    Naibaho, L., Nadeak, B., Sormin, E., & Juwita, C. P. (2021, June). The Analysis of Code-switching Integration Realization on Students Classroom Performance. In *2nd Annual Conference on blended learning, educational technology and Innovation (ACBLETI 2020)* (pp. 176-181). Atlantis Press.

45. Christian Religion Education as a Solution for Families to Face the Change in the Era of Revolution 4.0 (Atlantis-Press _ Advances in Social Science, Education and Humanities Research). **Proceeding published by Springer and indexed by WoS, 2021.**

    Rantung, D. A., & Naibaho, L. (2021, June). Christian Religion Education as a Solution for Families to Face the Change in the Era of Revolution 4.0. In *2nd Annual Conference on blended learning, educational technology and Innovation (ACBLETI 2020)* (pp. 260-265). Atlantis Press.

46. Effective School Management in Industrial Revolution Era 4.0 (Atlantis-Press _ Advances in Social Science, Education and Humanities Research). **Proceeding published by Springer and indexed by WoS, 2021.**

    Tyas, E. H., & Naibaho, L. (2021, June). Effective School Management in Industrial Revolution Era 4.0. In *2nd Annual*

Conference on blended learning, educational technology and Innovation (ACBLETI 2020) (pp. 212-216). Atlantis Press.

47. Students Perception on the Implementation of Higher-Educational Curriculum Based on Indonesian Qualification Framework at Postgraduate Program *Universitas Kristen Indonesia* (Atlantis-Press _ Advances in Social Science, Education and Humanities Research). **Proceeding published by Springer and indexed by WoS, 2021.**
    Simbolon, B. R., Sinaga, D., & Naibaho, L. (2021, June). Students Perception on the Implementation of Higher-Educational Curriculum Based on Indonesian Qualification Framework at Postgraduate Program Universitas Kristen Indonesia. In *2nd Annual Conference on blended learning, educational technology and Innovation (ACBLETI 2020)* (pp. 196-201). Atlantis Press.
48. Language Politeness (Atlantis-Press _ Advances in Social Science, Education and Humanities Research). **Proceeding published by Springer and indexed by WoS, 2021.**
    Simatupang, M. S., & Naibaho, L. (2021, June). Language Politeness. In *2nd Annual Conference on blended learning, educational technology and Innovation (ACBLETI 2020)* (pp. 166-171). Atlantis Press.
49. How Should the Classroom Be Managed During the Covid-19 Pandemic **(International Journal of Research-Granthaalayah, 9/5/2021 Page. 272–289.).**
50. Tyas, E. H., & Naibaho, L. (2021). How Should The Classroom Be Managed During The Covid-19 Pandemic?. *International Journal of Research-GRANTHAALAYAH*, *9*(5), 272-289
    The Analysis of Pre-Service EFL Teachers' Teaching Implementation (Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences, Vol 4/3/Page.3372-3381). **International Accredited Journal.**
    Naibaho, L. (2021). The Analysis of Pre-Service EFL Teachers' Teaching Implementation. *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences*, *4*(3), 3372-3381.
51. A Comparison of Female and Male English Lecturers' Working Performance and Competence: A Research-Based on Students'

Perspective **(Asian EFL Journal, SCOPUS INDEXED Journal _ Q1 _ 2021).**

52. The Effectiveness of Self-Regulated Strategy Development on Improving Students' Narrative Text Writing Achievement **(Tesol International Journal, SCOPUS INDEXED Journal _ Q1 _ 2021.**

53. The integration of mind mapping strategy on students' essay writing at Universities Kristen Indonesia. **National Accredited Journal**

Naibaho, L. (2022). The integration of mind mapping strategy on students' essay writing at universities kristen Indonesia. *JPPI (Jurnal Penelitian Pendidikan Indonesia), 8*(2), 320-328.

54. The analysis of students' reading and writing difficulties in learning english at Universitas Kristen Indonesia. **National Accredited Journal**

Naibaho, L. (2022). The analysis of students' reading and writing difficulties in learning english at universitas kristen indonesia. *Jurnal Konseling dan Pendidikan*, *10*(1), 157-166.

55. Exploring INQF-Based Curriculum in English Education Study Programs: Investigation of Curriculum Implementation. **National Accredited Journal**

Naibaho, L., & Silalahi, M. (2022). Exploring INQF-Based Curriculum in English Education Study Programs: Investigation of Curriculum Implementation. *Tarbawi: Jurnal Keilmuan Manajemen Pendidikan*, *8*(01), 113-120.

56. The Integration of Kohlberg Moral Development Theory with Education Character **(Technium Social Sciences Journal 31, 203-212) – International Journal.**

Nainggolan, M. M., & Naibaho, L. (2022). The Integration of Kohlberg Moral Development Theory with Education Character. *Technium Social Sciences Journal*, *31*, 203-212.

57. Community Learning Motivation to Join the School Package C Equivalently High School Education (Research at Foundation PKBM Imam Syafe'i, Bandung) – **National Accredited Journal**

Naibaho, L., Jura, D., & Afdaleni, A. (2022). Community Learning Motivation to Join the School Package C Equivalently High School Education (Research at Foundation PKBM Imam Syafe'i, Bandung). *ijd-demos*, *4*(1).

58. Development and implementation of Merdeka learning-merdeka campus at Christian university of Indonesia (International Journal of Advanced Educational Research 7 (1), 36-44) **International Journal**
Naibaho, L. (2022). Development and implementation of Merdeka learning-merdeka campus at Christian university of Indonesia. *International Journal of Advanced Educational Research*, *7*(1), 36-44.

59. English Teachers'professional Learning Community In Senior High Schools Students: Students'perspective On Its'implementation (Russian Journal of Education and Psychology 13 (3), 70-88). **International Journal Indexed by DOAJ.**
Naibaho L. English Teachers'professional Learning Community In Senior High Schools Students: Students'perspective On Its' implementation. Russian Journal of Education and Psychology. 2022 Jun 30;13(3):70-88.

60. The analysis of students' reading and writing difficulties in learning english at universitas kristen indonesia (Jurnal Konseling dan Pendidikan 10 (1), 157-166). **National Accredited Journal Sinta 2 and DOAJ indexed Journal.**
Naibaho L. The analysis of students' reading and writing difficulties in learning english at universitas kristen indonesia. Jurnal Konseling dan Pendidikan. 2022 Jul 21;10(1):157-66.

61. The Pre-Service Teachers'communication Ability Analysis At Teacher Training And Education Faculty (Indonesian EFL Journal 8 (2), 237-246). National **Accredited Journal Sinta 3 and DOAJ indexed Journal**.
Naibaho L. The Pre-Service Teachers'communication Ability Analysis At Teacher Training And Education Faculty. Indonesian EFL Journal. 2022 Jul 31;8(2):237-46.

62. The integration of mind mapping strategy on students' essay writing at universities kristen Indonesia (JPPI Jurnal Penelitian Pendidikan Indonesia) 8 (2), 320-328). **National Accredited Journal Sinta 2 and DOAJ indexed Journal.**
Naibaho L. The integration of mind mapping strategy on students' essay writing at universities kristen Indonesia. JPPI (Jurnal Penelitian Pendidikan Indonesia). 2022 Jun 16;8(2):320-8.

63. Pre-Service Efl Teachers'creative Thinking (Indonesian Journal of Learning and Instruction 5 (2). **National Accredited Journal Sinta 5 and DOAJ indexed Journal**.
    Naibaho L. Pre-Service Efl Teachers'creative Thinking. Indonesian Journal of Learning and Instruction. 2022 Nov 4;5(2).
64. The Role of Psycholinguistics in English Language Learning (Eximia 5 (1), 573-580).
    Naibaho L. The Role of Psycholinguistics in English Language Learning. Eximia. 2022 Nov 24;5(1):573-80. **International Journal**
65. Metode dan Peran Supervisi Pendidikan Meningkatkan Kualitas Pendidikan Agama Kristen Di SMA Negeri 1 Air Batu, Kabupaten Asahan Provinsi Sumatera Utara (Jurnal Christian Humaniora 6 (2). **National Journal.**
    Sidabutar N, Praing DI, Nadeak B, Naibaho L. Metode dan Peran Supervisi Pendidikan Meningkatkan Kualitas Pendidikan Agama Kristen Di SMA Negeri 1 Air Batu, Kabupaten Asahan Provinsi Sumatera Utara. Jurnal Christian Humaniora. 2022 Dec 1;6(2). **National Journal.**
66. Pemberdayaan KKG PAK dalam Peningkatan Kompetensi Profesional Guru PAK dalam Mengimplementasikan Kurikulum Merdeka (Jurnal Christian Humaniora 6 (2), 1-15).
    Purba S, Naibaho L, Rantung DA. Pemberdayaan KKG PAK dalam Peningkatan Kompetensi Profesional Guru PAK dalam Mengimplementasikan Kurikulum Merdeka. Jurnal Christian Humaniora. 2022 Dec 1;6(2):1-5. **National Journal.**

National and International Conferences

1. Developing English Teachers Professionalism to Meet the 21st Century Challenges **(Participants)**
2. Evaluating the Problematic of Character Education in Indonesia ***(Participants)***.
3. The 11th Asia TEFL International Conference on "English Across Asian Context: Opportunities and Challenges (Asia TEFL/**Presenter**)
4. The Implementation of 2013 Curriculum ***(Participants)***

5. The 2014 International Conference on Applied Linguistics and Language Education-ICALLE - De La Salle University, Manila Philippines **(Presenter)**
6. TheInternational Conference on Applied Linguistics and Language Education - ICALLE - De La Salle University, Manila Philippines **(Presenter)**
7. Barriers and Challenges of Christian Education and Its Solutions ***(participants)***
8. The Third European Conference on Education **(Presenter)**
9. Revitalizing the Role of Christian Youth in Realizing Indonesianity ***(participants)***
10. Mental Revolution in Education for Human Character Building ***(speaker)***
11. *Seminar and Workshop for Science Teacher of PSKD with the theme "Refreshing and Reframing Teacher Competencies" (Committee)* ***(speaker)***
12. *Seminar on Research Proposal Writing by FKIP UKI* ***(speaker)***
13. Teaching and Learning English in Indonesia "Future Trends and Approaches" ***(speaker)***
14. **I**naugural TESOL Indonesia International Conference ***(speaker)***
15. *LGBT in Scientists Perspective* ***(participants)***
16. TESOL Indonesia International Conference Edition December (Presenter)
17. The Asian EFL Journal International Conference on Research and Publication (Presenter) - Site Skill Training Campus, Clark.
18. The **3rd Women** in **TESOL** International Conference – Bali
19. 2nd The Asian EFL Journal International Conference on Research and Publication (Presenter) - Site Skill Training Campus, Clark. (Presenter)
20. National Seminar on "Sexuality Education begins in the Home and Porn Destroys Our Lives, Let's Fight it Together" Held by Teruna Muda Internasional School (***Keynote Speaker***)
21. Seminar on "The Family Intimacy" Held by Gereja HKBP Sutoyo Jakarta (***Invited Speaker***)

22. The 1st Jakarta International Conference on Science and Education, Held by Faculty of Education and Teacher Training, Universitas Kristen Indonesia (1stJICSE) (***Keynote Speaker***).
23. National Seminar "Writing an Article for Indexed Journal" Held by Faculty of Education and Teacher Training, Universitas Kristen Indonesia (***Invited Speaker***)
24. National Teacher Sharing on "Curriculum Modification" Held by Ikatan Alumni Sumatera Utara – Temu Kangen (***Invited Speaker***)
25. National Conference "Qualitative Research Method on Developing the Christian Theology and Christian Education on 4.0 Industrial Era, Held by Postgraduate Program Sekolah Tinggi Teologia Paulus, Medan (***Invited Speaker***).
26. National Seminar on Qualitative Research – Innovation of Social Research Method, Held by Indonesia Qualitative Researcher Association (***Invited Speaker***).
27. National Seminar on "Maintaining Teaching and Learning Spirit Admist the COVID-19 Pandemic" Held by Universitas Katolik Santo Thomas, Medan. ***(Invited Speaker)***
28. National Webinar on **"**Pelatihan Penelitian (Penulisan Proposal dan Metode Penelitian)**"** diselenggarakan oleh Program Magister Pendidikan Agama Kristen Universitas Kristen Indonesia pada 20 Maret 2021. ***(Invited Speaker)***
29. National Webinar on **"**Pelatihan Penulisan Jurnal**"** diselenggarakan oleh Program Magister Pendidikan Agama Kristen Universitas Kristen Indonesia pada 03 April 2021. ***(Invited Speaker).***
30. The 1st AEJ UKI SLA Research Conference *"English SLA in the Asian Context and Culture post Covid 19"*, held by Asian TESOL in Partnership with UKI on **April 23 – 25, 2021. *(Keynote Speaker).***

Visiting Lecturer/Scholar

1. Kazi Nazrul University, Department of Education, July 2020
2. STT. Theologi Paulus, Saturday, 12 September 2020
3. STT. Theologi Paulus, Saturday, 19 September 2020

Journal Editorial Board/Reviewer

1. Communita Servicio
2. International Journal of English Language Literature
3. Jurnal Dinamika Pendidikan
4. International Journal of Academic Library and Information Science
5. Lingua Cultura
6. Jurnal Eligible (LLdikti Wilayah III)
7. Jurnal Bilingual Universitas Simalungun
8. *Psychology Research and Behavior Management*

Trainings Program

1. Leadership Training – 1st
2. Writing Research Proposal Government Grant –
3. Leadership Training – 2nd
4. Writing Research Proposal Government Grant – 2nd
5. Thompson Reuters Indexed Proceeding Article Writing
6. Technical Guidance for Research Methodology
7. Training on Book Editor
8. Training on Social Mapping for CSR
9. Reaserch Collegium
10. International Qualitative Researcher Certification – Certified

Books

1. Becoming Great Hotilier (Neuro-Linguistics Programming for Hospitality): Formula NLP untuk Melayani Hingga Menangani Keluhan Tamu, Penerbit UKI Press, ISBN 978-623-7256-30-4, Year 2019.
2. The Power Creative Thinking and Imagination Suggestion on Writing: A Monograph Based on Research, Publisher: Widina Bhakti Persada Bandung, ISBN 978-623-6608-79-1, Year 2019
3. Moralitas Siswa dan Implikasinya dalam Pembelajaran Budi Pekerti (Kajian Teori Kohlberg dan Teori Lickona), Publisher: Widina Bhakti Persada Bandung, ISBN 978-623-6608-78-4, Year 2020
4. Psycholinguistics in Language Learning, Publisher: Widina Bhakti Persada Bandung a, ISBN 978-623-6092-32-3, Year 2021

5. Kepemimpinan & perilaku organisasi : konsep dan perkembangan, Publisher: Widina Bhakti Persada Bandung, ISBN 978-623-9325-54-1, Year 2020.
6. Philosophical Issues in Education: An Introduction, Publisher: Widina Bhakti Persada Bandung, ISBN 978-623-6457-52-8, Year 2021
7. Pengantar Penelitian Pendidikan, Widina Bhakti Persada Bandung, ISBN 978-6236457-45-0 Year, 2021

Modul
1. Pragmatics
2. Phonology
3. Morphology and syntax
4. English for Physics I
5. English for Biology
6. Introduction to General Linguistics
7. Psycholinguistics

Community Services
1. Achievement Motivation, 2) Personal Hygiene, and 3) Sex Education to Communities in Kepulauan Seribu, SD N 02 Pagi Pulau Kelapa
2. Socialization of the English Language Education Study Program
3. Counseling and training to parents about the use of educational methods, learning and skills for school children.
4. Socialization To The Teachers, Parents And Students About 1) Learning Motivation, 2) Self-Hygiene, And 3) Sex Education.
5. Socialization On English Education Department
6. Community Service "The 15^{th} Green Actions" *Kelurahan Cawang* Towards Green Environments, Independent And Without Drugs.
7. Workshop for The Students' Parents on The Use Of Education Method, Learning and Skills of The Students
8. Community Service on Teaching English to the Primary, Junior High School And Senior High School Students.
9. Fun English: Using Flash Cards And Realia For Young Learners At TK Gladi Siwi – Lubang Buaya Jakarta Timur
10. Fostering Marriage Resilience And Family Harmony With Theme “The Family Relationship And Intimacy

11. Church Social Service in GPIB Marturia Lampung
12. Citarum Harum
13. Rainwater Harvesting System (Water Harvesting) To Provide Raw Water And Clean Water in Bumi Dipasena Tulang Bawang Lampung
14. The Role of the Church and the Hkbp Family in Ending Crimes Against Women and Children

Institutional Occupation
1. Academic Advisor (2013 – present)
2. Head of University Curriculum Development (2014-2015)
3. Head of University Research (2016-present)

Certification
1. Certified Lecturer
2. Certified Book Editor
3. Certified International Qualitative Researcher

Achievement Appreciation
1. Certificate (Volunteer and Translator) - NGO Caritas Switzerland _ Based in Aceh Singkil.
2. Certificate (Outstanding Students Cumlaude with 3.69 GPA out of 4.00 scale) - State University of Medan.
3. Certificate and Charter (Outstanding Students _ Cumlaude with 4.00 GPA out of 4.00 scale) - Postgraduate Program _ the State University of Medan.
4. Certificate _ 3rd Winner on Articles Writing – APKASI
5. Certificate _ Top Ten on Scientific Article Writing - PT. Semen Indonesia
6. Research Grants - Government
7. Research Grants - Government

Working Experiences
1. PT. Alatan Indonesia _ **President Director** (July 2017 – Present)
2. Christian University of Indonesia _ (November 2016 – Present)
3. Christian University of Indonesia _ Lecturer (September 2012 – present)

3. Amik Universal _ Lecturer (Jun 2010 – April 2012)
4. Saint Paul Theologian Institution _Lecturer (Jun 2010 – April 2012)
5. IOM (International Organization for Migrant) _ English Consultant (January 2010-January 2012)
6. NGO-Caritas Switzerland _ English Teacher and Translator (June 2009 – December 2009).
7. BT / BS Bima Medan _ English Tutor (Jan 2006 – Des 2008)
8. PEEC (Prima Essential English Course) _ Teacher (Mate Teacher) (January 2002 – June 2005)

COMPILER'S BIBLIOGRAPHY

Lamhot Naibaho was born in Buluduri, November 18, 1985. He is the sixth of seven children, the son of Lamasi Naibaho and Sonti Aritonang. His father was a civil servant and his mother was a farmer. His education level starts from elementary school at 030404 Buluduri Elementary School, continues to junior high school at SMP Negeri 2 Laeparira and high school at SMA Negeri 1 Sidikalang. After that he continued his studies in 2005 at Medan State University in the Department of English Education, and graduated as a Cumlaude student in 2009. Then he continued his studies to a higher level at the Medan State University Postgraduate Program in 2010, and graduated as a student with the best achievement and the highest GPA (4.00) in 2012. At this time, he is taking a Doctoral Program at the State University of Jakarta and is completing his dissertation, entered in 2012. He started his career in 2002 while sitting in the first class chair High school as an English teacher at PEEC (Prima Essential English Course) in Sidikalang for three years. While sitting in the lecture chair, he was a guest at one of Paparon's Pizza for two years, then became an English tutor at the BIMA Learning Guidance in Medan for 2 years and also as a private English teacher. In the last semester of his undergraduate studies, he was chosen to become a translator and teacher assistant at NGO-Caritas Switzerland in Aceh Singkil. After that he was accepted as an English lecturer at Amik Universal and STT Paulus Medan and later became an English Consultant at IOM (International Organization for Migrants), and finally until now, he is a lecturer at the Indonesian Christian University in the English Study Program.