CHAPTER 4

# Spoken Word Recognition and Production

**Anne Cutler**

## I. INTRODUCTION

Most language behavior consists of speaking and listening. However, the recognition and production of spoken words have not always been central topics in psycholinguistic research. Considering the two fields together in one chapter reveals differences and similarities. Spoken-word recognition studies began in earnest in the 1970s, prompted on one hand by the development of laboratory tasks involving auditory presentation and on the other hand by the realization that the growing body of data on visual word recognition (see Seidenberg, Chapter 5, this volume) did not necessarily apply to listening, because of the temporal nature of speech signals. This recognition research has been, to a great extent, model driven. Spoken-word production research, on the other hand, has a longer history but a less intimate relationship with theory. Laboratory tasks for studying production are difficult to devise, so that much research addressed production *failure,* such as slips of the tongue. An unhappy result of this focus has been models of production that are largely determined by the characteristics of failed rather than successful operation of the processes modeled; moreover, the models have rarely prompted new research. Only in recent years have models and laboratory studies of successful production become widely available. Despite these differences, theoretical questions in recognition and produc-

*Speech, Language, and Communication*

tion are similar, and central to both has been the question of autonomy versus interaction in levels of processing. As this chapter describes, the balance of evidence indicates a similar answer for both fields.

## II. SPOKEN WORD RECOGNITION

### A. Methodology

The simplest methods for studying spoken word recognition measure recognizability per se, for example, of words or sentences under noise masks (e.g., Savin, 1963) or of filtered or truncated ("gated"; Ellis, Derbyshire, & Joseph, 1971) speech signals. None of these tasks reflects the time course of recognition.

Among tasks that do attempt to tap processing on-line are those of lexical decision, deciding whether or not a stimulus is a real word (see Seidenberg, Chapter 5, this volume); as this merely requires affirming that recognition has occurred, it has a certain ecological validity. In the auditory form of this task, however, subjects cannot respond before the end of a word in case it becomes a nonword at the last opportunity (*recognitious* rather than *recognition,* for instance); thus response time (RT) varies with word length (Bradley & Forster, 1987). Also, subjects can be reluctant to reject nonwords, perhaps reflecting all listeners' practice in guessing from unclear speech input. Variants of lexical decision are "word-spotting" (Cutler & Norris, 1988), in which a largely nonsense input contains occasional real words and a single positive response to a real word replaces lexical decision's choice response, and "phoneme-triggered lexical decision" (Blank, 1980), in which subjects listen for a specified initial phoneme and then decide whether it begins a word or nonword. Both can be used with continuous speech input. Speeded repetition of words ("auditory naming"; Whalen, 1991) or continuous speech ("shadowing"; Marslen-Wilson, 1985) may also be used to study recognition, but these tasks have the disadvantage of confounding recognition and production processes.

In phonetic categorization, principally used to study segmental perception (see Nygaard and Pisoni, Chapter 3, this volume), the stimuli are *unnatural* tokens, along a continuum formed by some acoustic parameter between one natural speech sound and another. For word recognition studies the tokens may be embedded in words versus nonwords (e.g., Ganong, 1980).

In dual-task experiments, RT to perform the secondary task can reflect momentary changes in difficulty of the primary task. Many such tasks require subjects to monitor incoming speech for a prespecified target, which can be part of the speech stream itself (a word, a syllable, a phoneme), something wrong with the speech (e.g., a mispronunciation; Cole, 1973), or an extraneous signal (e.g., a click; Ladefoged & Broadbent, 1960). In phoneme monitoring (Foss, 1969), subjects detect a target phoneme (usually in word-initial position). The task must be performed at a phonemic rather than an acoustic match level, since the acoustic patterns of phonemes differ with context. Phoneme monitoring has featured in many studies of prelexical processing, as has syllable monitoring (e.g., Mills, 1980b), in which the target is a sequence such as /ba/. Rhyme monitoring (detection of a word rhyming with the target; e.g., Seidenberg & Tanenhaus, 1979) has been used to study phonological characteristics of lexical entries. In word monitoring (e.g., Marslen-Wilson & Tyler, 1980), specification of the target involves recognition of the very word to which a response is later made; responses are therefore subject to repetition priming, that is, facilitation of recognition on a second occurrence (e.g., Slowiaczek, Nusbaum, & Pisoni, 1987).

Another indirect measure is cross-modal priming (Swinney, Onifer, Prather, & Hirschkowitz, 1979), in which listeners perform a lexical decision to a visually presented string while simultaneously listening to words or sentences. Response time to the visual words is measured as a function of their associative relationship to the spoken words. Facilitatory effects, and the time at which they appear relative to presentation of the spoken word, reflect lexical processing of the latter. This task has proved particularly useful for studying patterns of activation in the lexicon during word recognition.

## B. Issues

### 1. The Input to the Lexicon

Lexical access from spectra (LAFS; Klatt, 1979) is the only recognition model in which lexical hypotheses are generated directly from a spectral representation of input. In this model, the lexicon consists of a tree of possible sequences, in which the end of every word is connected to the beginning of every other word, with phonological rules supplying the appropriate allophone; the result is a decoding network for phonetic sequences. Each state transition in this network is a mininetwork of spectral templates. Lexical recognition consists of finding the best match between the input sequence of spectral templates and paths through the network.

LAFS is neither an implemented engineering system nor a psychological model specified in terms that make predictions about human performance. Its main advantage is that it offers a way to preserve useful low-level information that is often lost in the process of making all-or-none phonemic decisions. However, it incorporates some psychologically implausible features, such as the redundancy of having similar word boundary transitions represented for every possible word pair separately. Problems of variability (Nygaard and Pisoni, Chapter 3; this volume) eventually proved insuperable for actual implementation of LAFS (Klatt, 1989). The model requires a spectral distance metric for the computation of matching scores for each input spectrum with the stored spectral templates; but the spectral variability of real speech input defeated efforts to devise such a metric.

Other models of spoken word recognition assume that speech input is translated into a more abstract representation for contact with the lexicon. Obvious candidates for such representations have been the units of linguistic analysis, and of these, the phoneme, by definition the smallest unit into which speech can be sequentially decomposed, has been most popular (Foss & Gernsbacher, 1983; Marslen-Wilson & Welsh, 1978; Pisoni & Luce, 1987). However, other units have been proposed, including units above the phoneme level such as syllables (Mehler, Dommergues, Frauenfelder, & Segui, 1981), demisyllables (i.e., a vowel plus a preceding syllabic onset or following coda; Fujimura & Lovins, 1978; Samuel, 1989), diphones (i.e., speech segments affected by two consecutive phonemes; Klatt, 1979; Marcus, 1981) or stress units (Grosjean & Gee, 1987), and feature units below the phoneme level (McClelland & Elman, 1986b; Stevens, 1986; Stevens, Manuel, Shattuck-Hufnagel, & Liu, 1992) (see Nygaard and Pisoni, Chapter 3, this volume, for further discussion).

The issue of prelexical representations connects with the issue of lexical segmentation of continuous speech. Understanding speech requires recognition of individual words (more exactly: lexically represented units), since the number of possible complete utterances is infinite. Yet speech signals do not necessarily contain robust and reliable cues to word boundaries. Several classes of solution have been proposed for the problem of segmenting continuous speech into words. One is that words are recognized in sequential order, and that the point of onset of a word is identified by virtue of successful recognition of the preceding word (Cole & Jakimik, 1978; Marslen-Wilson & Welsh, 1978). Another is that alternative, possibly overlapping, word candidates compete for recognition (this is the mechanism embodied in TRACE: McClelland & Elman, 1986b). The third class of solution proposes an explicit segmentation procedure (e.g., Cutler & Norris, 1988). Explicit segmentation may presuppose units of segmentation at the prelexical level.

Experimental evidence supports all three types of proposal to some extent. Sequential recognition, however, is poorly suited to vocabulary structure, as described in Section IIB2, and competition is fully compatible with explicit segmentation. Evidence exists for explicit segmentation (1) into syllables: listeners detect target strings such as *ba* or *bal* more rapidly when the strings correspond exactly to a syllable of a heard word than when they constitute more or less than a syllable (Mehler et al., 1981; Zwitserlood, Schriefers, Lahiri, & van Donselaar, 1993); and (2) at stress unit boundaries:

recognition of real words embedded in nonsense bisyllables is inhibited if the word spans a boundary between two strong syllables (i.e., two syllables containing full vowels), but is not if the word spans a boundary between a strong and a weak syllable, since only the former is a stress unit boundary (Cutler & Norris, 1988).

A striking outcome of the explicit segmentation research is that segmentation units appear to be language specific. The evidence for syllables reported above comes from French and Dutch. Evidence of syllabic segmentation has also been observed in Spanish (Bradley, Sánchez-Casas, & García-Albea, 1993) and Catalan (Sebastian-Gallés, Dupoux, Segui, & Mehler, 1992). Other tasks confirm the robustness of syllabic segmentation in French (Kolinsky, Morais & Cluytens, 1995; Segui, Frauenfelder, & Mehler, 1981). However, target detection does not show syllabic segmentation in English (Cutler, Mehler, Norris & Segui, 1986) or in Japanese (Otake, Hatano, Cutler, & Mehler, 1993; Cutler & Otake, 1994). Cutler and Norris's (1988) observation that segmentation in English is stress based, by contrast, is supported by patterns of word boundary misperceptions (Cutler & Butterfield, 1992) and by evidence of activation of monosyllabic words embedded as strong syllables in longer words (e.g., *bone,* in *trombone;* Shillcock, 1990).

Stress-based segmentation in English and syllabic segmentation in French are similar in that in both cases segmentation procedures appear to reflect the basic rhythmic structure of the language; this parallel led Otake et al. (1993) to investigate segmentation in Japanese, the rhythm of which is based on a subsyllabic unit, the mora. Otake et al.'s results (see also Cutler & Otake, 1994) supported moraic segmentation in Japanese.

## 2. Constraints of Temporal Structure

Spoken words are temporal signals. The temporal nature of spoken word recognition is the basis of the Cohort Model of Marslen-Wilson and Welsh (1978), which models word recognition via bottom-up analysis of the input combined with top-down influence from knowledge-driven constraints. Each stored lexical representation is an active element, which responds to both bottom-up and top-down information. The initial portion of a spoken word activates all words beginning in that way. Thus /s/ activates, for example, *sad, several, spelling, psychology.* As more information arrives, processing elements that do not match the input drop out of this initial cohort of potential words. If the next phoneme is /p/, for example, the cohort is reduced to words beginning /sp/; an incoming /ɪ/ reduces it still further, to *spinach, spirit, spill,* and so on until only one candidate word remains. This may occur before the end of the word. Because no words other than *spigot* begin /spɪg/, those four phonemes suffice to reduce the cohort to one word.

The model capitalizes on the temporal nature of speech in that initial portions of words affect the recognition process first. Recognition is efficient in that a word can be recognized by that point at which it becomes unique from all other words in the language (the "uniqueness point"), which may (as in *spigot*) precede the word's end.

In support of the uniqueness point concept, Marslen-Wilson (1984) reported correlations between phoneme detection RT and the distance of target phonemes from uniqueness points, and between the point at which listeners confidently recognized gated words and the words' uniqueness points. Similarly, nonwords that are the first part of real words (e.g. [krɛd] as in *credit*) are rejected more slowly than are nonwords that cannot become words (e.g. [krɛn]; Taft, 1986); mispronunciations later in words are detected faster and more accurately than changes in initial sounds (Cole, 1973; Marslen-Wilson & Welsh, 1978). In support of the concept of an activated cohort, Marslen-Wilson (1987) found that presentation of either *captain* or *captive* initially activated words related to both (e.g., *ship, guard*), suggesting that both potential words were themselves activated; at the end of either word, however, only associates of the word actually presented were facilitated. Marslen-Wilson and Zwitserlood (1989) reported that words were not primed by other words with similar ending portions; thus recognition of *battle* was primed by *batter,* but not by *cattle*. However, Connine, Blasko, and Titone (1993) found that nonwords that differed from real words by only one or two phonological features of the initial phoneme primed associates of their base words (i.e.,*deacher* primed *school*).

Other evidence also suggests that words are not necessarily recognized as early as possible, and that word-final information may be fully processed. Gated words are usually recognized after their uniqueness point, sometimes only after their acoustic offset (Bard, Shillcock, & Altmann, 1988; Grosjean, 1985). Lexical decision RTs to words with the same uniqueness point and same number of phonemes after the uniqueness point (e.g., *difficult, diffident*) differ as a function of frequency; and listeners' latency to reject nonwords with equal uniqueness points is affected by what follows this point (e.g., *rhythlic,* which has the same ending as *rhythmic,* is rejected more slowly than *rhythlen,* although both become nonwords at the /l/; Taft & Hambly, 1986; see also Goodman & Huttenlocher, 1988). Words in sentences can be identified either from word-initial or word-final information, and both types of information interact with higher-level contextual information (Salasoo & Pisoni, 1985); truncated words can be recognized from either initial or final fragments (Nooteboom, 1981); and beginnings and ends of synthesized words are equally effective recognition prompts when the rest of the word is noise masked (Nooteboom & van der Vlugt, 1988). Identification of spoken words is facilitated by prior presentation of rhyming items (Milberg, Blumstein & Dworetzky, 1988) and, under noise masking, by prior presentation of words overlapping either from onset or offset; recognition of *flock* is facilitated as much by *stock* or *block* as by *flap* or *flop* (Slowiaczek et al., 1987).

Indeed, the uniqueness point concept may have limited application, since most short words (in English) become unique only at their end (*starve, start*), or even after it (*star* could continue as *starve, start, starling,* etc.; Luce, 1986), and most words (even content words) in conversational speech are monosyllabic (Cutler & Carter, 1987). Counterproposals consider instead the overall similarity of a word to other words (Luce, Pisoni, & Goldinger, 1990; Marcus & Frauenfelder, 1985). All-or-none rejection of candidate words due to early mismatch is also inconsistent with detection of word-initial mispronunciations (Cole, 1973), with word-initial phoneme restoration (Warren, 1970), and with recognition of improbable utterances such as puns (Norris, 1982). Marslen-Wilson (1987) revised the cohort model to cope with these problems. The later version of the model assumes that minor mismatches only downgrade a candidate word's activation rather than remove it entirely from the cohort. However, the model still gives highest priority to word-initial information.

## 3. The Limits of Interaction

### *a. Interactive Models of Spoken Word Recognition*

The Logogen Model (Morton, 1969) is a model of how sources of information combine in word recognition, originally designed for visual recognition, and for data just on visual recognition thresholds at that. But it has been treated as a general model of word recognition by many researchers. "Logogens," in this model, are active feature counters associated with individual lexical items, which tot up incoming positive evidence in their individual favor. Each logogen has a threshold, at which it outputs a recognition response. Not all logogens are equal: those for high-frequency words, for instance, start with higher resting levels than those for low-frequency words. As the input may come not only from auditory features but also from higher-level sources such as syntactic and semantic analysis of the preceding speech context, the logogen model is interactive. Probably because it is less fully specified than more recent models, the logogen model has not played a major role in recent years.

In the first version of the Cohort Model, higher-level context could reduce the cohort: words inconsistent with the developing context drop out. Thus, strong contextual constraints would produce even earlier recognition. Consistent with this, recognition of gated words showed facilitatory effects of both syntactic and semantic constraints (Tyler, 1984; Tyler & Wessels, 1983); and in a shadowing task, predictable mispronounced words were fluently restored to correct pronunciation, as if listeners had not noticed the errors (Marslen-Wilson, 1975). Further, Marslen-Wilson and Welsh (1978) found that the tendency of mispronunciations late in a word to be more often fluently restored than word-initial mispronunciations was significantly stronger for predictable than for unpredictable words. In the revised cohort model, however, higher-level context does not act directly on cohort membership. A set of potential candidates generated by bottom-up input is evaluated and integrated with context; information from prior context flows down only to this integrative stage, not as far as lexical selection. By this latter change the cohort model is no longer strictly interactive, but essentially an application to auditory word recognition of Norris's (1986) noninteractive model of context effects in visual word recognition, and similar to proposals about spoken word recognition made by Luce et al. (1990).

The case for interactivity in spoken word recognition is put most strongly by the connectionist model TRACE. A miniversion of the model (McClelland & Elman, 1986a) takes real speech input, but is of limited scope (one speaker and a vocabulary of nine CV syllables). The better-known version (McClelland & Elman, 1986b) finesses front-end analysis problems in the interest of explicating the higher-level capabilities; the input is an explicit representation of acoustic features. Both versions of the model embody interactive activation. On a series of levels are processing elements (nodes), each with (1) a current activation value, (2) a resting level, toward which its activation decays over time, and (3) a threshold, over which its activation spreads to connected nodes. Bidirectional connections exist between nodes within each level and at adjacent levels. TRACE mimics the temporal nature of speech by representing each node separately in each one of successive time slices.

Connections between nodes at the same level can be excitatory (e.g., connection at the phoneme level between /t/ and phonemes in adjacent time slices that can precede or follow /t/, such as vowels), or inhibitory (e.g., connection between /t/ and other phonemes in the same time slice). Between-level connections (e.g., from /t/ to all features at the feature level, below, which occur in /t/, or to all words containing /t/ at the word level, above) are always facilitatory. Connections are assigned weights (positive for excitatory connections, negative for inhibitory). Interactive activation is the process of constantly adjusting connection weights, and hence the activation of nodes, according to signals received from other nodes, on the same, lower, or higher levels. The input is a pattern of activation of feature nodes, presented sequentially across time slices. Activation spreads upward from the feature level, while within the feature level inhibition is exerted toward features incompatible with those activated by the input. As higher-level nodes are excited from the bottom up, their increasing activation leads them to exert influence from the top down, which in turn increases activation from the bottom up, and so on. Recognition occurs (at any output level) when a node's activation reaches some arbitrary level. A decision can in principle be made at any time, the current most highly activated node being deemed the output (McClelland & Elman, 1986b, p. 20). Alternatively, output could occur at a preset time rather than at a preset activation level.

An advantage of TRACE is its exploitation of variability in cues to phonemes. For instance, the fact that acoustic cues for stop consonants vary with the identity of the following vowel is represented in the miniversion of TRACE by connections that allow feature nodes for vowels to adjust the weights of feature nodes for prevocalic consonants. TRACE's major limitation, however, is that it deals ineffectively with the temporal nature of speech; multiplication of the entire network across time slices is a highly implausible proposal. Subsequent connectionist models of speech recognition exploited techniques better adapted to deal with temporal variation, such as simple recurrent networks (Elman, 1990; Norris, 1990). According to Klatt (1989), current connectionist models also provide only limited and implausible solutions to the problems of variability in speech.

Note that simultaneous consideration of input and higher-level information does not presuppose interactivity. In Massaro's (1987) Fuzzy Logical Model of Speech Perception (FLMP), multiple sources of information about speech input are simultaneously but independently evaluated in a continuous manner, and are integrated and compared as to the relative support they offer for stored prototypes. It is called a fuzzy logical model because continuous truth values are assigned to each source of information as a result of the evaluation process. The FLMP offers a way of integrating information from input and context without positing interaction: the information from each source is evaluated independently, and one does not modify the processing of the other. Massaro (1989) has argued that in signal detection theory terms, his model predicts top-down effects only on bias (β), not sensitivity (d′), whereas TRACE, which allows activation from higher levels of processing to feed back to and alter lower levels of processing, predicts that top-down effects will alter sensitivity; Massaro reported results consistent with the FLMP's predictions.

### *b. Lexical and Phonetic Processing*

Evidence of word/nonword differences in the earliest stages of speech perception would argue strongly for interaction. Three main lines of research have addressed this issue. The first, using the phonetic categorization task, began with Ganong (1980), who asked listeners to identify CVC-initial consonants varying along a voicing continuum, one choice consistent with a word and the other with a nonword (e.g., *teak–deak* vs. *teep–deep*). The crossover point of listeners' identification functions shifted, such that more choices were lexically consistent. The lexical shift was largest in the ambiguous region of the continuum, which Ganong interpreted as a top-down effect, arguing that response bias would have affected the entire continuum equally.

Fox (1984), however, found that fast categorization responses were unaffected by lexicality (also see Miller & Dexter, 1988). Fox proposed that phonetic and lexical representations are computed in parallel; activation of lexical representations takes longer than phonetic computation, but once lexical information is available, it dominates, with the result that only later responses are lexically biased. Connine and Clifton (1987) replicated Fox's RT result, but argued that if the lexical effect was indeed a postperceptual response bias, its effects should mimic those of another bias such as monetary payoff; they did not, renewing support for top-down interpretation of lexical effects on categorization. Also, TRACE successfully simulated (McClelland & Elman, 1986b) both Ganong's (1980) and Fox's (1984) findings, as well as effects of phonological legality in categorization (Massaro & Cohen, 1983).

An influential study by Elman and McClelland (1988) capitalized on one type of compensation for coarticulation, namely identification of alveolar and velar stops after dental versus palatal fricatives. Lip rounding for the palatal [ʃ] elongates the vocal tract, while lip retraction for the dental [s] shortens it, resulting in changes in the articulation of any following stop; but listeners compensate for these: an ambiguous sound halfway between [t] and [k] tends to be reported as [t] if it occurs after [ʃ], but as [k] if it occurs after [s] (Mann & Repp, 1981). Elman and McClelland appended an ambiguous fricative between [ʃ] and [s] to words normally ending in [ʃ] (*foolish*) and [s] (*Christmas*), and had listeners judge an ambiguous stop consonant between [t] and [k], presented in word-initial position immediately following the ambiguous fricative. A compensation shift appeared in listeners' judgments after the ambiguous as well as after the unambiguous fricative. The ambiguous fricative provided no acoustic reason for compensation; Elman and McClelland therefore interpreted the shift in terms of top-down lexical influence inducing biased judgments about acoustic structure at the phonetic decision level.

In phonetic categorization experiments, of course, many stimuli are by definition unnatural, that is, not what speakers ever produce; this makes the segmental perception task both difficult and unrepresentative of natural recognition. Also, the same words are responded to repeatedly, which may preactivate lexical representations and hence make influence of lexical information more likely. Recent results with the task suggest that lexical effects are fragile and rather dependent on unnaturalness of the stimuli. Burton, Baum, and Blumstein (1989) found that lexical effects on [d]–[t] categorizations disappeared when the stimuli covaried burst amplitude with VOT (and were thus more like natural speech). Likewise, McQueen (1991) found no lexical effects in word-final [s]–[ʃ] categorizations with high-quality stimuli, though lexical effects appeared when the same stimuli were degraded by filtering. (Note that McQueen's finding of small and unreliable lexical effects on word-final phoneme categorization also contradicts TRACE, which predicts stronger lexical effects word-finally than word-initially; as word information accumulates, top-down feedback should strengthen.) Compare these inconsistent lexical status effects with effects of speaking rate, actual or inferred (e.g., Miller, 1981a, 1981b; Miller & Liberman, 1979). A continuum varying formant transition duration from [b] (short) to [w] (long) produces more [b] judgments before a long [a], more [w] judgments before a short [a], indicating that listeners compensate for (inferred) rate of articulation by computing transition duration in relation to stimulus duration (Miller & Liberman, 1979). This effect appears with severely filtered phonetic context (Gordon, 1988), suggesting that it operates at an early, coarse-grained, level of phonetic processing; moreover, the effect cannot be removed by, for instance, requiring fast responses (Miller & Dexter, 1988; Miller, Green, & Schermer, 1984). Thus, the rate effect seems to be automatic, and robust in a way that the effects of lexical status are not. The presence of lexical status effects, as Pitt and Samuel (1993) concluded from a comprehensive meta-analysis of the phonetic categorization studies, depends heavily both on position of the phoneme in the stimulus item, and on the phonetic contrast involved. In all, the pattern of variability suggests that the phonetic categorization task may not be the best way to study lexical–phonetic relationships.

In phoneme restoration studies (Warren, 1970), a phoneme is excised and replaced by noise; listeners report hearing noise simultaneous with an intact speech signal (something that they have often experienced). Their judgments of the location of the noise are imperfect, but the limits of illusory continuity correspond roughly to average word duration (Bashford, Meyers, Brubaker, & Warren, 1988; Bashford & Warren, 1987), suggesting that once lexical information is available, it supplants lower-level information. Samuel (1981a) compared listeners' judgments when noise actually replaced a phoneme (as it did in Warren's stimuli) or was overlaid on a phoneme (as listeners reported). Discriminability ($d'$) of replaced versus overlaid stimuli was significantly worse in real words than in nonwords, and in high-frequency words than in low-frequency words (although Samuel, 1981b, failed to replicate the frequency effect). Discriminability was also worse when the distorted word was preceded by the same word, undistorted. Bias ($\beta$) showed no effects either of lexical status or of frequency. Samuel (1987) similarly found that words with early uniqueness points were less discriminable (at least for distortions late in the word) than words with late uniqueness points; and words with more than one completion (e.g., *-egion*) were less discriminable than words with a unique completion (*-esion*). Samuel interpreted his results in terms of top-down lexical influence upon phonetic decisions: previously activated or otherwise easily identifiable words are more speedily accessed and feedback from their lexical representations overrides the low-level discrimination, while nonwords remain discriminable because no lexical representations exercise influence. Real words are, however, no more detectable under noise than nonwords (Repp & Frost, 1988).

The third line of research uses phoneme-monitoring and asks whether responses are pre- or postlexical. Postlexical responding is suggested by findings that targets are detected faster on predictable than on unpredictable words (Eimas & Nygaard, 1992; Mehler & Segui, 1987; Morton & Long, 1976), targets may be detected faster on words than on nonwords (P. Rubin, Turvey, & van Gelder, 1976), and word targets are detected faster than phoneme targets on the same words (Foss & Swinney, 1973). Likewise, if subjects monitor for targets occurring anywhere in a word, not just word-initially, targets are detected faster on high- versus low-frequency words (Segui & Frauenfelder, 1986) and on contextually primed versus unprimed words (Frauenfelder & Segui, 1989). On the other hand, Foss and Blank (1980; also see Foss, Harwood, & Blank, 1980) found no RT effects of word–nonword status or frequency of occurrence of the target-bearing word; and when subjects monitored for word-initial phonemes only no effects of word frequency (Segui & Frauenfelder, 1986) or contextual priming appeared (Frauenfelder & Segui, 1989). Repeated findings that the level of match between target specification and response item affects RT in this task (Cutler, Butterfield, & Williams, 1987; Dijkstra, Schreuder, & Frauenfelder, 1989; Healy & Cutting, 1976; McNeill & Lindig, 1973; Mills, 1980a, 1980b; Swinney & Prather, 1980; Whalen, 1984, 1991) are consistent with prelexical responses, as is the interference effect of phonemes phonologically similar to the target (Newman & Dell, 1978). Newman and Dell argued that a phonological representation constructed postlexically should allow a simple yes–no decision on match of input to target; but RT in their study showed a gradient of interference as a function of number of shared phonological features. Further, this interference effect is as strong in predictable as in unpredictable words (Dell & Newman, 1980; Stemberger, Elman, & Haden, 1985).

The apparently conflicting phoneme-monitoring findings can be accounted for by assuming that responses can be made either pre- or postlexically, as in the Race Model of phoneme detection proposed by Cutler and Norris (1979; also see Foss & Blank, 1980; Newman & Dell, 1978). In this model, parallel processes compute an explicit phonemic representation and access the lexicon, with a phoneme detection response possible on the basis of whichever representation is available first; the model contrasts with TRACE, which allows phonemic responses only from the phonemic level. Whether or not lexical effects on phoneme detection RT are observed depends on characteristics of the materials and of the subjects' task (Cutler, Mehler, Norris, & Segui, 1987; Foss & Gernsbacher, 1983); these can affect the level of response at which subjects' attention is focused and, in turn, which process wins the phoneme response race. TRACE successfully simulated (McClelland & Elman, 1986b) both Foss and Gernsbacher's (1983) failure to find lexical effects on word-initial phoneme detection, and Marslen-Wilson's (1980) finding that phoneme detection *did* show lexical influence late in a word. Counter to TRACE's predictions, however, is a finding by Frauenfelder, Segui, and Dijkstra (1990) on phoneme detection in nonwords. In TRACE, incoming phonetic information consistent with a particular word facilitates all phonemes in that word, which in turn leads to inhibition of phonemes not in the word. Thus TRACE predicts that RT to [t] in *vocabutary* will be slow (compared with RT to [t] in, say, *socabutary*), because incoming information will be consistent with the word *vocabulary*, which contains no [t]. The prediction was not borne out: RTs for the two types of nonword did not differ. This finding is consistent with the Race Model, which assumes a prelexical focus of attention with nonword input.

Attentional accounts have similarly been offered for RT differences in word versus phoneme monitoring (Brunner & Pisoni, 1982), for the disappearance of top-down effects under speeded response conditions in phonetic categorization (Miller et al., 1984), and for changes in the relative weight of cues to phonemic identity (Gordon, Eberhardt, & Rueckl, 1993). In phonemic restoration, Samuel (1991; Samuel & Ressler, 1986) elicited significantly more accurate discriminations of replaced versus overlaid phonemes by telling subjects which phoneme to attend to, suggesting that subjects can switch attention from the lexical level (which normally mediates the restoration illusion) to the phonetic level. In phoneme monitoring, listeners can be induced to attend selectively to position in the word (Pitt & Samuel, 1990) or syllable (Pallier, Sebastian-Gallés, Felguera, Christophe, & Mehler, 1993). Eimas, Hornstein, and Payton (1990) found that listeners making a phoneme choice response (e.g., [b] or [p]) normally attended to the phonetic level, but addition of a secondary task requiring a lexical judgment resulted in attention switching to the lexical level, such that lexical effects upon RT appeared. In sentence contexts, however, the same items did not produce reliable lexical effects with or without a secondary task (Eimas & Nygaard, 1992; cf. Foss & Blank, 1980; Foss et al., 1980), suggesting that attention switching is less easy in coherent contexts than in word lists. This in turn suggests that in phoneme monitoring, as in phonetic categorization, lexical effects may be highly dependent on those characteristics of the experimental situation that least resemble natural recognition conditions.

### *c. Sentence and Word Processing*

How far should top-down influences from syntactic and semantic context extend? No current models propose direct interaction between nonadjacent processing levels, and effects of sentence context on phonetic processing seem better explained as response bias than as perceptual effects. Phonetic categorizations can be shifted by contextual probability, so that a sound between [b] and [p] can be judged [b] in *hot water for the —ath* but [p] in *jog along the —ath* (Miller et al., 1984); however, the shift disappears in speeded responses, suggesting that it is not a mandatory perceptual effect. Similarly, Connine (1987) found that sentence context affected phonetic categorization RT for stimuli at the continuum end points but not at the category boundary, again indicative of response bias rather than an alteration of perceptual decisions; and Samuel (1981a) found that the bias measure in phonemic restoration (which showed no lexical status or frequency effects) *was* influenced by sentence context, but the discriminability measure was not. Both Samuel and Connine argued that sentential context does not constrain phonetic processing, though they held that it may affect it indirectly by constraining lexical processing.

Consistent with constraints of sentence context on lexical processing are faster word monitoring in syntactically and semantically acceptable contexts than in semantically anomalous or ungrammatical contexts (Marslen-Wilson, Brown, & Tyler, 1988; Marslen-Wilson & Tyler, 1980) and earlier recognition of gated words in constraining semantic and syntactic contexts (Grosjean, 1980; Tyler & Wessels, 1983). Prior presentation of associates of the target-bearing word facilitates phoneme-monitoring RT when the words are in the same sentence (though such effects are weak in lists: Blank & Foss, 1978; Foss, 1982; Foss & Ross, 1983). However, Foss explained this phoneme-monitoring facilitation in terms of ease of integration of the word meaning into a representation of the sentence meaning, and Marslen-Wilson et al. (1988) accounted similarly for the word-monitoring finding. Connine, Blasko, and Hall (1991) provide converging evidence of rapid and efficient integration mechanisms. Context effects in gating experiments may result from guessing (given that, after all, guessing on the basis of partial information is what the subject has to do). Thus, none of these findings entail that availability of lexical candidates is constrained by the sentence context. Pertinent evidence on this issue comes from the case of homophony.

When a listener hears an ambiguous string such as [wik], sentence context could constrain lexical processing such that only the contextually appropriate sense of the word is retrieved; alternatively, if lexical processing is independent of syntactic and semantic processing, all senses may be activated, and the appropriate one chosen by reference to the context. Using the cross-modal priming task, Swinney (1979) found that homophones such as *bug* primed words related to both their senses, even when prior context was consistent with only one of the senses; although after just a few syllables, only the contextually appropriate sense was active. Facilitation of words related to both senses occurs even when one reading of the ambiguous word is far more frequent than the other(s): *scale* primes both *weight* and *fish* (Onifer & Swinney, 1981), and even when there are word class differences and syntactic context permits only one reading: *week/weak* primes both *month* and *strong* (Lucas, 1987).

Similarly, a spoken homophone facilitates naming of a visually presented word related to either one of its senses, irrespective of form class; *flower* is named more rapidly after either *She held the rose* or *They all rose* than after control sentences, and again the multiple-sense activation is short lived (Tanenhaus, Leiman, & Seidenberg, 1979). The same methodology produces automatic activation of multiple senses for noun–verb as well as noun–noun ambiguities (Seidenberg, Tanenhaus, Leiman, & Bienkowski, 1982). Lackner and Garrett (1972) presented listeners with two competing messages, and required them to attend to one and to paraphrase it. Speech in the unattended channel (which subjects could not report) resolved ambiguities in the attended utterances; subjects' paraphrases reflected either sense, depending on the available disambiguation, again suggesting availability of all senses.

All these results suggest momentary simultaneous activation of all senses of a homophone, irrespective of relative frequency or contextual probability, with postaccess decision processes rapidly and efficiently selecting an appropriate sense and discarding inappropriate sense(s). This would imply that lexical access is *not* subject to top-down influence from syntactic and semantic processing. Note that Seidenberg et al. (1982) found that strong associates in the preceding context *could* constrain one sense of a homophone; but they interpreted this as an effect of lexical association rather than of syntactic or semantic context. Similarly, completely restrictive syntactic contexts fail to prevent activation of multiple senses of noun–verb homophones (e.g., *sink;* Oden & Spira, 1983; Tanenhaus & Donnenwerth-Nolan, 1984), although Oden and Spira found a greater degree of facilitation for contextually appropriate than for contextually inappropriate senses. Tabossi (1988a; also see Tabossi, Colombo, & Job, 1987) found that strongly constraining contexts could lead to only one sense being activated if that particular sense was highly dominant (e.g., the *harbor* sense of *port* in *ships in the port*). But again, these contexts effectively primed the relevant sense via occurrence of a related word, contexts that constrained one sense without priming it (e.g., *The man had to be at the port*) produced facilitation for all senses. Thus, a within-lexical explanation is again as viable as a top-down account. Exhaustive activation of the senses of an ambiguous word does *not* occur when the word is not focused in the sentence (Blutner & Sommer, 1988), suggesting that the kind of activation of related words that the cross-modal priming task taps may occur only when attention is directed to the part of the sentence containing the priming word.

Unambiguous words can also have senses of varying relevance to particular contexts. All attributes may be momentarily activated when a word is heard, irrespective of their relative dominance and of their contextual appropriateness (Whitney, McKay, Kellas, & Emerson, 1985); shortly after word offset, however, attributes that are dominant and/or contextually appropriate are still active, but contextually inappropriate nondominant attributes are not. Similarly, central properties of unambiguous words (e.g., that *ice* is *cold*) are activated irrespective of contextual appropriateness, but peripheral properties (e.g., that *ice* is *slippery*) are activated only when appropriate (Greenspan, 1986; this latter experiment had a delay between context and target, and thus resembles Whitney et al.'s postoffset condition).

Tabossi (1988b) found that sentence contexts *could* constrain activation of different aspects of an unambiguous word's meaning; HARD was primed after *The strong blow didn't crack the diamond,* but not after *The jeweler polished the diamond.* However, Williams (1988) found that even in neutral contexts there could be a failure of priming, so that TABLE is not primed after *The boy put the chair. . . .* (although *chair* reliably primes TABLE out of context). Williams argued that the relations between prime and target here involved schematic world knowledge rather than lexical associations; he proposed that activation of background knowledge during construction of the semantic representation of a sentence determined the kind of attribute to which the listener's attention would be directed. Only certain kinds of semantic relation, however, are subject to such attentional constraint; other kinds of relation produce truly context-independent priming effects. This argument suggests that the scope for sentence context effects on lexical processing may be limited to certain kinds of relationship between a word and its context. Even then, context effects are, as argued above, apparently located at a relatively late stage of processing. Zwitserlood (1989) showed that both contextually appropriate and inappropriate words are momentarily activated before sufficient input is available to select between them; thus context does not constrain the availability of lexical candidates per se.

## 4. Competition for Lexical Selection

First proposed in TRACE, the notion that candidate words actively compete for recognition is achieving wide popularity. Evidence of activation of words embedded within or across other words (Cluff & Luce, 1990; Shillcock, 1990; Tabossi, Burani & Scott, 1995), or of simultaneous activation of partially overlapping words (Goldinger, Luce, & Pisoni, 1989; Goldinger, Luce, Pisoni, & Marcario, 1992; Gow & Gordon, 1995; Marslen-Wilson, 1990; Zwitserlood, 1989) is consistent with the competition notion, but does not entail it. Inhibition of recognition as a function of the existence of competitors provides direct evidence. Taft (1986) observed that nonwords that form part of real words are hard to reject. Direct evidence of competition between word candidates comes from the finding that *mess* is harder to recognize when preceded by the syllable [də] (as in *domestic*) than when preceded by, say, [nə] (McQueen, Norris, & Cutler, 1994), and from the finding that recognition is affected by the number of other words potentially accounting for a portion of the input (Norris, McQueen & Cutler, 1995; Vroomen & de Gelder, 1995).

Analysis of patterns of competition depends crucially on precise knowledge of vocabulary structure. Studies of lexical structure have been revolutionized in recent years by the availability of computerized dictionaries; it is now easy to analyze the composition of the vocabulary in many languages, and arguments based on analyses of lexical databases have come to play an important role in theorizing about spoken word recognition (e.g., Cutler & Carter, 1987; Luce, 1986; Marcus & Frauenfelder, 1985). It should be noted, however, that substantial corpora of spoken language, and the estimates of spoken word frequency that could be derived from them, are still lacking; such spoken word frequency counts as do exist (e.g., G. D. A. Brown, 1984; Howes, 1966) are, for practical reasons, small in scale compared to written frequency counts.

Recent modeling initiatives have been designed to exploit vocabulary structure effects. In the Neighborhood Activation Model (Luce et al., 1990), word decision units are directly activated by acoustic/phonetic input, with no top-down information from context playing a role at this stage. Once activated, the decision units monitor both input information and the level of activity in the system as a whole; their activity levels are also heavily biased by frequency. Thus in this model, the probability of recognizing a spoken word is a function both of the word's own frequency and of the number and frequency of similar words in the language: high-frequency words with few, low-frequency neighbors are recognized rapidly and accurately, while low-frequency words with many high-frequency neighbors are much harder to recognize (Goldinger et al., 1989; Luce et al., 1990). So far, the model has been implemented only for monosyllabic words (indeed, CVC words only). The model is similar to the Cohort model in determining initial activation by bottom-up influence only, but differs from it both in the relative importance of word-initial information (in NAM, information pertaining to any part of a word is equally important) and in the central role of frequency. It is similar to TRACE in assuming that recognition is crucially dependent on the pattern of activity in the system as a whole.

Specifically incorporating the notion of competition is SHORTLIST (Norris, 1991, 1994), a hybrid connectionist model that has strong similarities with both the revised Cohort model and TRACE, but is burdened neither with the Cohort model's overdependence on initial information nor with TRACE's multiplication of the lexical network across time slices. Furthermore, unlike TRACE, it is a strictly autonomous model. SHORTLIST's main feature is the separation of recognition into two distinct stages. In the first, bottom-up information alone determines the set of potential word candidates compatible with the input (the "short list"); the set may include candidates that overlap in the input. The initial stage can be implemented as a simple recurrent network (Norris, 1990), although it has been simulated as a dictionary search by Norris (1994). The short-listed candidates are then wired into a small interactive activation network, containing only as many connections as are needed for the particular set of words being processed; these words then compete for recognition. Because the competition stage is limited to a small candidate set, the model can be implemented with a realistic vocabulary of tens of thousands of words (Norris, 1994).

As noted earlier, the notion of competition provides a potential solution to the problem of word recognition in continuous speech; even without word boundary information, competing words may effectively divide up the input amongst themselves. However, competition can also co-exist with explicit segmentation. When interword competition and stress-based segmentation are compared in the same experiment, independent evidence appears for both (McQueen et al., 1994).

## III. SPOKEN WORD PRODUCTION

### A. Methodology

The laboratory study of word production raises formidable problems; ensuring that a particular word is produced may subvert the spontaneous production process, for example, by making the word available to the subject's *recognition* system prior to its production. Tasks in which subjects output on command a previously learned sequence (e.g., Sternberg, Knoll, Monsell, & Wright, 1988) or paired associate (Meyer, 1990) may preserve later stages of the production process but preempt earlier stages. Presumably because of this difficulty with designing laboratory studies, much research on production has exploited naturalistic data: slips of the tongue, tip-of-the-tongue (TOT) states, pausing, and hesitation patterns. Word production in particular has been investigated via slips and TOT, in other words, primarily via instances of processing failure. Even though well-controlled laboratory conditions have been devised to collect slips (Baars, Motley, & MacKay, 1975), and articulatory failure in tongue twisters (Shattuck-Hufnagel, 1992) and tip-of-the-tongue states (Kohn et al., 1987), the possibility that the processes inferred from production failures are not those of normal production still bedevils this field.

Word production has also been studied via the picture-naming task, which has a long history in psychology (see Glaser, 1992); in particular, picture naming under conditions of interference from simultaneously presented words (e.g., Lupker, 1979) has allowed inferences about connections within the lexicon (cf. the cross-modal priming task for recognition). In recent versions of this task subjects hear a word and see a picture, and the asynchrony of presentation of these two stimuli may be varied; the dependent variable may be a response to the picture (e.g., Meyer & Schriefers, 1991) or to the spoken word (e.g., Levelt et al., 1991). Again, drawing inferences about word production from these tasks requires assumptions about word recognition as well. Picture-naming experiments of any kind are further limited in that they are applicable only to the study of words that can be depicted: concrete nouns, plus a few readily encodable action verbs (e.g., Kempen & Huijbers, 1983). There are still large areas of vocabulary for which no experimental investigation technique has yet been devised.

## B. Issues

### 1. Stages in Word Production

As Levelt (1989) points out, the conceptual formulation stages of production have attracted more attention from outside psychology (e.g., from AI approaches to language generation) than from within it. There is naturalistic and empirical work on propositional and syntactic formulation (see Bock, Chapter 6, this volume), but virtually none that specifically addresses the input to lexical processing.

Picture-naming studies that partially address the issue include Huttenlocher and Kubicek's (1983) observation that picture naming can be primed by prior presentation of pictures of related objects; they argued that, however, because the size of the priming effect is the same in picture naming as in object recognition without naming, what is primed is recognition of the picture only, and not retrieval of the appropriate name from the lexicon. Flores d'Arcais and Schreuder (1987) found that the naming of pictures (e.g., of a guitar) was facilitated both by prior presentations of other pictures having functional relations to the target picture (e.g., an accordion) and by pictures of objects physically similar to the target (e.g., a tennis racket); they argued for separate involvement of functional and physical aspects of a picture in specification of the relevant lexical input. As in Huttenlocher and Kubicek's study, however, their effects may be located in picture processing rather than in lexical retrieval.

Whatever the precise nature of the input, lexical access in production involves a mapping from semantic to phonological form; a lexical representation can act as a transcoding device that accomplishes this mapping (see Fromkin, 1971, for such a proposal, and see Bock, Chapter 6, this volume). However, separation of semantic/syntactic and phonological information even within the lexical system for production has been assumed by many models. In its simplest form (e.g., Butterworth, 1989) this proposal assumes that each word is associated with a conceptual node, which is accessed by input from message generation, and that this node is in turn connected to a separate phonological node that sends the output to the subsequent stage of the production process. The influential language production model of Garrett (1975) laid greatest emphasis on separation of two distinct levels of processing, corresponding closely to a distinction between two types of words, open-class (or lexical, or content) and closed-class (or grammatical, or function) words. At the "functional" level, only open-class items and their grammatical relations are specified. At the "positional" level all words in an utterance are specified, including inflectional endings that specify grammatical relations. The former level is held to contain no aspects of phonological structure (or at least none that are capable of exercising independent effect at that level); the latter, although still held to be a stage prior to full phonetic specification, "involves certain aspects of sound structure." Although this proposal is not formulated in terms of lexical processing, it is compatible with a two-stage model of lexical processing in which syntactic and phonological forms are separately accessed.

A two-stage proposal for lexical access in production was made by Kempen and Huijbers (1983), and the word production part of Levelt's (1989) model of utterance production is likewise two stage. The conceptual nodes accessed in the first stage of lexical processing these authors call *lemmas*, the phonological nodes, *word forms*. Although Levelt (1989) is neutral as to whether lemmas and word forms are accessed together or separately, he subsequently (Levelt, 1992) proposed that the two stages are temporally sequential.

Corresponding to the notion of two stages, there are modeling efforts devoted principally to one or to the other. For the mapping from concept to lexicon, Oldfield (1966; Oldfield & Wingfield, 1964) proposed successive choice between binary semantic feature alternatives of increasing specificity (+/−*living*, +/−*animate*, +/−*human*, etc.). As Levelt (1989) points out, however, in any such decompositional system the semantic mapping that enables production of, for example, *terrier* will also suffice for production of *dog* (and indeed for more general terms at every level of the semantic hierarchy; *animal*, *creature*, etc.). Levelt terms this "the hyperonym problem." Semantic decomposition within lexical representations is proposed by Bierwisch and Schreuder (1992). In their model, the mapping from concepts to lemmas is mediated by the construction of an utterance semantic form that may allow alternative instantiations of a conceptual intention as sequences of lemmas; a verbalization function chooses the instantiation that will be mapped to the lexicon. The hyperonym problem is assumed to be solved by restricting lemma access to the closest match to the input semantic form. Roelofs (1992) proposes a spreading activation model of lemma access, in which lemma nodes receive activation from associated concept nodes, and the node with the highest activation level is "selected"; the lemma nodes are linked in turn to word-form nodes, so that selection of a lemma node presumably acts as the criterial stimulus for activation of the associated word-form node. Spreading activation solutions could in principle solve the hyperonym problem by ensuring that a specific input (*terrier*) will produce greater activation in more specific lemmas than in more general ones (*dog*). However, the problem does not arise in Roelofs's model because of the one-to-one mapping from concepts to lemmas. A decompositional process of concept access would, of course, merely shift the problem to the concept-node level. Roelofs (1993) argues, however, that decompositional acquisition of concepts is compatible with later chunking of the conceptual components to a unitary representation, such that concept access in production short-circuits the process of concept construction from semantic primitives.

The mapping from semantic to phonological structure in language production is reflected in slips at each level. In semantic misselection, a semantically related word may substitute for the target word (*the two contemporary* [T: *adjacent*] *buildings*), or two semantically related words may be simultaneously chosen and the output form may be a blend of both (*science fiction bookstops* [T: *stores/shops*]). In the latter case the two components of the blend are equally good alternative selections (in the particular context: Bierwisch, 1981). In phonological misselection, substitution errors may produce words with similarity of sound but not of meaning: *participate* for *precipitate* (Fay & Cutler, 1977). However, such pairs apparently do not blend (errors such as *partipitate* are not reported). Misselections that are only semantic or only phonological, and differing constraints on semantic versus phonological errors, are consistent with separation of semantic from phonological information in lexical access.

For word-form retrieval, Shattuck-Hufnagel (1979, 1983, 1986) proposes a model in which a suprasegmental framework, consisting of prosodically labeled syllable frames, is separately specified (presumably by extralexical considerations), and segmental components of the phonological forms of words are copied into it by a 'scan-copy" device. The latter selects and copies the segments in serial order. As each segment is selected it is checked off; the copied representation is constantly monitored. Shattuck-Hufnagel's model is primarily intended to account for patterns of phoneme slips: phoneme anticipations and exchanges occur when the scanner selects the wrong segment; perseverations are failures of check-off. Like other error-based models, Shattuck-Hufnagel's assumes that word forms consist of phoneme representations ordered in syllabic structure. Slips in the pronunciation of words preserve syllable structure: syllable-initial phonemes exchange with one another, syllable-final with one another, and so on. Patterns of error in tongue twisters reveal the same pattern (Shattuck-Hufnagel, 1992). However, slips at differing levels of syllable structure are possible: thus an initial consonant cluster can move as a unit (e.g., exchange with another syllable onset: *cledge hippers*), or one element of the cluster can move alone (*Sprench feaker*). Phonemes can be misordered, as in the above examples, or misselected (omitted, added, substituted: *overwelling* for *overwhelming, osposed* for *opposed, neasely* for *neatly*). Such slips involve movement or selection of whole phonemes, not of the phonological features in terms of which phonemes can be described (Shattuck-Hufnagel & Klatt, 1979). Slips are more likely when a sound occurs twice (e.g., *a lilting willy*); the two occurrences should have the same role in syllable structure (Dell, 1984). Slips are more likely between, and more likely to create, two words with the same pattern of syllable structure(s) (Stemberger, 1990).

Phoneme similarity can precipitate error: the likelihood of error increases with increasing similarity between phonemes, both in natural (Shattuck-Hufnagel, 1986; Shattuck-Hufnagel & Klatt, 1979) and in experimentally elicited slips (Levitt & Healy, 1985). Levitt and Healy also found that frequent phonemes tend to displace infrequent ones, and argued that the former have stronger output representations than the latter. However, no such asymmetry was observed in Shattuck-Hufnagel and Klatt's (1979) natural slips corpus; nor did they find that markedness was a relevant variable in error frequency (but see Stemberger, 1991). Similarity effects may occur in mapping a phonological to an articulatory code.

Speakers producing a learned word on command initiate production more rapidly if they are sure of the initial syllable of the word to be spoken, even more rapidly if they are sure of the initial two syllables, but knowledge of noninitial syllables alone cannot speed production (Meyer, 1990). Speakers naming a picture produce their response faster if just as they see the picture they hear a word with the same onset as the name, or just after they see the picture they hear a word with the same offset as the name (Meyer & Schriefers, 1991). Both results suggest that word encoding produces onsets before offsets rather than whole words at once.

## 2. Interaction between Stages of Production

### *a. Interactive Models of Word Production*

Word production has been less systematically modeled than recognition. Morton's logogen model (1969), as a general theory of lexical representation and retrieval, is applicable equally to recognition and production. In the case of production, the evidence that raises a logogen's activation level will normally be conceptual input from the cognitive system, and the output once a logogen has reached threshold is an (unspecified) phonological code. Just as logogens are sensitive to semantics as well as phonology in recognition, they are sensitive to phonological as well as semantic information in production.

The model proposed by Dell (1986, 1988), in which lexical processing within sentence production is carried out by spreading activation within a network, contrasts most obviously with the noninteractive two-stage models of Levelt (1989) and others. Semantic, morphological, and phonological aspects of any lexical item's structure are represented separately within this network, but there is no division of the network to reflect this separation, with the result that activation can flow in either direction within any combination of these aspects of representation. Similar proposals for a lexical network have been made by Stemberger (1985) and MacKay (1987).

### *b. Sentence and Word Processing*

Interaction between lexical and contextual processing in production would require the demonstration of feedback from lexical information to subsequent prelexical decisions about utterance content. Bierwisch and Schreuder (1992) propose that the matching of lexical representations to conceptual input involves distributed processing, but their model does not include feedback from lexicon to the input. Levelt and Maassen (1981) argued on the basis of a finding that word accessibility (as defined by naming latency) did not affect order of mention that there is *no* such feedback. Harley (1984) discusses speech errors in which recently activated words intrude into an utterance, for example, *I've eaten* (T:*read*) *all my library books,* spoken by a hungry person preparing to eat. But such errors do not imply that activation of, for example, food-related words in the lexicon has resulted in a change to the speaker's *plan* for the utterance; rather, as Garrett (1988) argues, they can be explained as high-level blend errors, in which multiple utterance plans, generated in parallel, may merge.

It has been argued that syntactic formulation precedes lexical selection (Fromkin, 1971), follows it (Bierwisch & Schreuder, 1992), or operates in parallel to it (Bock, 1982). Bock (1986) argued that phonological information cannot feed back to influence word availability during syntactic formulation. In her experiment, speakers were asked to describe pictures, for example, of a church being struck by lightning; prior presentation of a prime word tended to determine choice of subject noun when the prime was semantically related to one of the concepts in the picture (e.g., after the prime *thunder* speakers were more likely to start with *lightning*), but phonologically related primes (e.g.,*frightening*) had no such effect. This result is compatible with models of lexical processing in which semantic and phonological information become available separately, and only the former is available to the syntactic formulator.

However, Bock (1987) found, using the same technique, that prior presentation of a prime word having maximal overlap from onset with a target word (e.g., *plan* for *plant*) inhibited choice of the target as first noun in a picture description. Bock argued that prior presentation of a close phonological neighbor makes word forms temporarily less accessible, which in turn prompts the syntactic formulator to choose a structure in which access of that form is delayed. Levelt and Maassen (1981) found that although word accessibility did not affect order of mention, it did affect syntactic formulation in that where nouns were less accessible, picture descriptions were spoken more slowly and contained less syntactic reduction (e.g., coordinate reduction). These findings could suggest feedback from word-form selection to syntactic formulation, but Levelt (1989) accommodates them within an autonomous framework by proposing that relative word-form inaccessibility causes an ongoing utterance production to grind to a halt, and the resulting (unspecified) distress signal prompts an earlier process (here, syntactic formulation) to start again with an alternative plan.

### *c. Lexical and Phonetic Processing*

Natural phoneme misordering errors (Dell & Reich, 1981) and experimentally elicited phoneme slips (Baars et al., 1975) tend to result in real words (rather than nonwords) more often than chance would predict. Also, many word substitution errors show simultaneous semantic and phonological relationships (e.g., *typhoid;* T: *thyroid;* Aitchison & Straf, 1981). Baars et al. proposed a prearticulatory output monitor to explain the asymmetry, but Dell and Reich argued for simultaneous accessibility of semantic and phonological aspects of words (as proposed by Dell, 1986). As described above, this proposal conflicts with the autonomous two-stage models. Schriefers, Meyer, and Levelt (1990) asked subjects to name simple pictures (e.g., of a finger); naming time was slowed if subjects heard a semantically related word (e.g., *toe*) just before the picture appeared, but it was speeded if they heard a phonologically related word (e.g., *finch*) just *after* the picture appeared. Levelt et al. (1991) further adapted this paradigm such that the dependent variable became response time to make a lexical decision to the spoken word, presented just after the picture. When the spoken word was the name of the picture itself or was either semantically or phonologically related to this name, responses were inhibited in comparison to responses to a word unrelated to the picture. However, responses to words that were phonologically related to the picture's semantic associates (e.g., in the above example, *tone,* similar to *toe*) were not inhibited. Dell's (1986) spreading activation model would predict that any activation would be simultaneously semantic and phonological, counter to Levelt et al.'s results. Dell and O'Seaghdha (1991, 1992) propose a refinement of Dell's model, which brings it closer to the two-stage model; the access of semantic and of phonological information is modeled as separate stages subject to only limited interaction (presumed necessary to explain the speech error effects).

## IV. RECOGNITION AND PRODUCTION

Apart from the obvious difference that the mapping between sound and meaning proceeds in opposite directions, what comparisons are appropriate between the processes of recognition and production of spoken words? Are the processes interdependent? Do they, for instance, draw on a single lexicon, or two? This issue has arisen regularly, despite the fact that most models address only one or the other system. In the logogen model, an exception to this separation, simultaneous sensitivity to both phonological and semantic input suggests a unitary lexical representation for each word; but later developments of the theory (Morton & Patterson, 1980) divided the lexicon into input versus output logogens, with the latter sensitive only to semantic information. Support for this view came from a finding by Shallice, McLeod, and Lewis (1985) that auditory name detection and reading words aloud can be performed simultaneously without cross-task interference. On the other hand, word substitution errors resembling the intended word in sound but not in meaning led Fay and Cutler (1977) to argue for a single lexical system, in which such production errors arise by misselection of a neighbor in an organization determined by the needs of recognition. This concluding section considers the similarities and differences in the structure of the recognition and production systems (via consideration, first, of some research issues that have arisen in both fields) (and see Bock, Chapter 6, this volume).

### A. Some Common Issues in Recognition and Production Research

#### 1. Frequency of Occurrence

Relative frequency of occurrence affects word recognition (see also Seidenberg, Chapter 5, this volume); for instance, accuracy of report of words in noise rises with frequency (Howes, 1957; Luce et al., 1990; Savin, 1963). However, because listeners produce many high-frequency erroneous responses, Savin (1963) ascribed this result to response bias rather than within-lexical effects; the same argument was made by Luce et al. (1990) because frequency effects on word identification were distinct from effects of neighborhood density (how many phonetically similar words could be confused with the target word), and by Connine, Titone, and Wang (1993) because frequency effects in a phonetic categorization task could be induced by manipulating overall list frequency. Luce et al. proposed that frequency effects occur at a late stage of word recognition, in which decision units, that can be biased by frequency, select among an initial set of candidates consistent with bottom-up information. (The revised cohort model uses a similar frequency mechanism: Marslen-Wilson, 1987.)

High-frequency words are recognized faster in auditory lexical decision (Connine, Mullenix, Shernoff, & Yelen, 1990; Dupoux & Mehler, 1990; Taft & Hambly, 1986; Tyler, Marslen-Wilson, Rentoul, & Hannay, 1988) but frequency effects are stronger with mono- than with polysyllabic words (Bradley & Forster, 1987); in phoneme monitoring, too, frequency effects appear only with targets on monosyllabic words (Dupoux & Mehler, 1990; Foss & Blank, 1980). Monitoring tasks are assumed to tap prelexical or access stages of processing. Where tasks tap later stages of processing, frequency effects are stronger, for example, in rhyme monitoring (McQueen, 1993). Frequency effects in gating disappear in constraining context (Grosjean & Itzler, 1984). The pattern of findings is thus consistent with models that place frequency effects at final decision rather than initial activation stages of lexical processing.

In production, frequency effects are stronger in picture naming than in naming a written word (Huttenlocher & Kubicek, 1983; Oldfield & Wingfield, 1965), but are not accounted for by object recognition time (Wingfield, 1968). Frequent words are less subject to phonological error than infrequent words (Dell, 1990; Stemberger & MacWhinney, 1986b). Word substitutions involving semantic associates tend to replace low-frequency with higher-frequency words (Levelt, 1989). TOT states are more common on low- than on high-frequency words (possibly better estimated by subjective than objective frequency; R. Brown & McNeill, 1966). The facilitatory effects of frequency suggested by these findings are instantiated in the logogen model (Morton, 1969) by differing levels of resting activation, in spreading-activation models by different weights on connections (e.g., MacKay, 1987), or by the number of connections to representations of possible contexts (Dell, 1990). However, frequency effects in homophones (*great/grate*) are determined by combined frequency rather than individual sense frequency (Dell, 1990; Jescheniak & Levelt, 1994), suggesting that they are located at word-form rather than lemma level. This in turn suggests that frequency effects in recognition and production have an ordering similarity (they arise late in word processing) but do not share a common location in a unitary system.

## 2. Morphological Structure

When two words exchange places in a speech error they frequently strand their inflectional affixes (e.g., *looking for boozes in his insect*); the affixes accommodate to their new stems (the plural inflection that would have been pronounced [s] on *insect* becomes [əz] on *booze*). This suggests separate representation of stem and inflectional affix in word production (Garrett, 1988; Stemberger & Lewis, 1986). Production of past-tense forms given infinitive verbs takes longer for irregular (*taught*) than for regular forms (*talked;* MacKay, 1976). Stems that end in the same phoneme as their inflection (e.g., *yield/yielded, doze/dozes*) are more subject to inflectional omission errors than stems ending in different phonemes (*grab/grabbed, change/changes;* Stemberger & MacWhinney, 1986a). Recognition of regularly inflected, but not irregularly inflected, forms facilitates later recognition of the stem (Kempley & Morton, 1982). Listeners perform same/different judgments faster on stems than on inflections (Jarvella & Meijers, 1983). These observations are all compatible with storage of inflected words in their base form. However, Stemberger and MacWhinney (1986b) argue that common inflected forms are stored as wholes, because they undergo error less often than uncommon inflected forms.

Inflectional affixes may therefore be stripped off in recognition; Taft, Hambly, and Kinoshita (1986) proposed the same for derivational prefixes, because RT to reject a nonword is longer if the nonword begins with a real prefix (e.g.,*dejouse* vs. *tejouse*). Against this suggestion are findings that RTs to detect mispronunciations in second versus first syllables are unaffected by whether or not the first syllable is a prefix (Cole & Jakimik, 1980), and that lexical decision RTs do not suggest that processing of prefixed words (e.g., *intention*) proceeds via processing of unrelated unprefixed words embedded in them (*tension;* Schriefers, Zwitserlood, & Roelofs, 1991; Taft, 1988; Tyler et al., 1988). A morphologically complex word such as *permit* is recognized more rapidly if subjects have just heard another word with the same stem, such as *submit,* but prior presentation of an affix does not have the same effect (Emmorey, 1989). Emmorey argued for a model of the lexicon in which words with the same stem are linked to one another, but having the same affix does not involve such links. Morphological structure is reflected in subjects' TOT guesses (D. C. Rubin, 1975); and when lexical stress is misplaced in production (e.g., *econOmists*), it virtually always falls on a syllable that does bear stress in a word morphologically related to the target (e.g., *economic;* Cutler, 1980), suggesting that lexical representations of morphologically related words are linked in production, too.

Garrett (1988) has argued that the separation of syntactic affixes from stems in production implies autonomy of syntactic from conceptual processing; the same argument was made for spoken word recognition by Katz, Boyce, Goldstein, and Lukatela (1987), who found that recognition time for inflectionally related Serbo-Croatian nouns was affected by syntactic form but not by frequency (which, it will be recalled, seems to be a late effect in recognition).

### 3. Phonological Structure

Proposals of the form of phonological representation in the recognition lexicon vary from the highly concrete, with contextual variation explicitly represented (e.g., Elman & Zipser, 1988) to the highly abstract, with all predictable information unspecified (Lahiri & Marslen-Wilson, 1991). In production, some sequencing errors apparently involve movement of elements postulated for underlying but not surface phonological representations (e.g., *swin and swaig;* T: *swing and sway;* Fromkin, 1973; the underlying representation of [ŋ] is hypothesized to be [ng]); others involve replacement of elements that would be underspecified in an abstract representation by elements that would be specified (Stemberger, 1991). Both these findings suggest that phonological representations for production may be in abstract form.

The phonological information in lexical entries includes syllable structure, in particular a division between onset and rime: speakers find it easier to play novel language games that require onset–rime division of words than games that require other divisions (Treiman, 1983, 1986); word blends most often involve division of component words between onset and rime (MacKay, 1972). Likewise, syllabic onsets exhibit perceptual integrity (Cutler, Butterfield, & Williams, 1987). Although the syllable per se is not a perceptual unit in English, it is in French and Dutch (see Section IIB1), and converging evidence for a role for the syllable in word production also comes from Dutch (Meyer, 1990). Similarly, studies in Japanese suggest that morae but not syllables play a role in recognition (Otake et al., 1993; Cutler & Otake, 1994) and production (Kubozono, 1989).

Lexical representation of word prosody in production has been proposed on the basis of TOT guesses that maintain stress pattern (R. Brown & McNeill, 1966), slips that maintain everything *but* stress pattern (Cutler, 1980), and stress priming effects in picture naming (Levelt, 1993). In recognition, lexical effects in phonetic categorization can be mediated by stress (Connine, Clifton, & Cutler, 1987) and by lexical tone (Fox & Unkefer, 1985). However, the purely prosodic correlates of stress pattern in English do not play a role in prelexical processing (Cutler, 1986), although misstressed words are hard to recognize if vowel quality alters (Bond & Small, 1983; Cutler & Clifton, 1984; Slowiaczek, 1990).

### B. The Architecture of the System

The evidence on overlap between the production and recognition systems is inconclusive. If there are to be shared resources, the constraints of the two processes entail that the sharing must be at a central level; however, while the evidence outlined above is compatible with a shared-resource account, it also does not rule out separation of the two systems. There is clearly room for innovative research on this issue.

Similarly, no conclusive answer can as yet be given to the central issue in both recognition and production research: is the basic architecture of the processing system autonomous/modular or interactive (see Seidenberg, Chapter 5, this volume)? Models of both types exist that can account for most existing evidence. However, there are reasons to prefer, at this point, an autonomous solution. First, a unidirectional flow of information presupposes a simpler architecture than bidirectional connections; if the explanatory power of each type of model is equivalent, then a choice can at least be made on the basis of simplicity. Second, autonomous models make clearer predictions about the sources of information accessible at each stage of processing, and hence are (at least in principle) more amenable to experimental test. Third, we have seen that many of the findings that have been regarded as favoring interactionist positions are far from robust. In production, phonological formulation is impervious to influence from semantic factors, and lexical accessibility does not strongly influence utterance formulation [in fact, in the only study that produced results apparently indicative of such influence (Bock, 1987), the significant priming effect was just 3% different from the chance prediction; in Levelt and Maassen's (1981) comparable study, an effect of similar size was not significant). In recognition, the tasks that have been used to investigate the relationship between phonetic and lexical processing have proved to be subject to task-specific effects. In particular, the phonetic categorization task, which requires repeated responses to the same stimuli, seems unsuited to investigation of normal recognition. Of the other tasks, phonetic restoration has produced evidence that on balance favors an interactive account, while the evidence from phoneme monitoring more strongly supports an autonomous position. Likewise, the literature offers little evidence for strong determination of lexical recognition by higher-level context. Some a priori less-likely interpretations may be effectively ruled out by context; but context effects may simply be indications that integrative processes in recognition are extremely efficient. With efficient integration, does the system actually need top-down processes, which, by excluding improbable interpretations, might actually mislead?

## Acknowledgments

Thanks for helpful comments on earlier versions of this text to Peter Eimas, David Fay, Pim Levelt, James McQueen, Joanne Miller, and Dennis Norris, and further thanks to Sally Butterfield, Francois Grosjean, David Pisoni, and Arthur Samuel.

## References

Aitchison, J., & Straf, M. (1981). Lexical storage and retrieval: A developing skill? *Linguistics, 19,* 751–795.

Baars, B., Motley, M., & MacKay, D. G. (1975). Output editing for lexical status from artificially elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 14,* 382–391.

Bard, E. G., Shillcock, R. C., & Altmann, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: Effects of subsequent context. *Perception & Psychophysics, 44*, 395–408.

Bashford, J. A., Meyers, M. D., Brubaker, B. S., & Warren, R. M. (1988). Illusory continuity of interrupted speech: Speech rate determines durational limits. *Journal of the Acoustical Society of America, 84*, 1635–1638.

Bashford, J. A., & Warren, R. M. (1987). Multiple phonemic restorations follow the rules for auditory induction. *Perception & Psychophysics, 42*, 114–121.

Bierwisch, M. (1981). Linguistics and language error. *Linguistics, 19*, 583–626.

Bierwisch, M., & Schreuder, R. (1992). From concepts to lexical items. *Cognition, 42*, 23–60.

Blank, M. A. (1980). Measuring lexical access during sentence processing. *Perception & Psychophysics, 28*, 1–8.

Blank, M. A., & Foss, D. J. (1978). Semantic facilitation and lexical access during sentence processing. *Memory & Cognition, 6*, 644–652.

Blutner, R., & Sommer, R. (1988). Sentence processing and lexical access: The influence of the focus-identifying task. *Journal of Memory and Language, 27*, 359–367.

Bock, J. K. (1982). Toward a cognitive psychology of syntax: Information processing contributions to sentence formulation. *Psychological Review, 89*, 1–47.

Bock, J. K. (1986). Meaning, sound and syntax: Lexical priming in sentence production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 12*, 575–586.

Bock, J. K. (1987). An effect of the accessibility of word forms on sentence structures. *Journal of Memory and Language, 26*, 119–137.

Bond, Z. S., & Small, L. H. (1983). Voicing, vowel and stress mispronunciations in continuous speech. *Perception & Psychophysics, 34*, 470–474.

Bradley, D. C., & Forster, K. I. (1987). A reader's view of listening. *Cognition, 25*, 103–134.

Bradley, D. C., Sánchez-Casas, R. M., & García-Albea, J. E. (1993). The status of the syllable in the perception of Spanish and English. *Language and Cognitive Processes, 8*, 197–233.

Brown, G. D. A. (1984). A frequency count of 190,000 words in the London-Lund Corpus of English Conversation. *Behavior Research Methods, Instrumentation and Computers, 16*, 502–532.

Brown, R., & McNeill, D. (1966). The "tip of the tongue" phenomenon. *Journal of Verbal Learning and Verbal Behavior, 5*, 325–337.

Brunner, H., & Pisoni, D. B. (1982). Some effects of perceptual load on spoken text comprehension. *Journal of Verbal Learning and Verbal Behavior, 21*, 186–195.

Burton, M. W., Baum, S. R., & Blumstein, S. E. (1989). Lexical effects on the phonetic categorization of speech: The role of acoustic structure. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 567-575.

Butterworth, B. (1989). Lexical access in speech production. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 108–135). Cambridge, MA: MIT Press.

Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 551–563.

Cole, R. A. (1973). Listening for mispronunciations: A measure of what we hear during speech. *Perception & Psychophysics, 11*, 153–156.

Cole, R. A., & Jakimik, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of information processing* (pp. 67–116). London: Academic Press.

Cole, R. A., & Jakimik, J. (1980). How are syllables used to recognize words? *Journal of the Acoustical Society of America, 67*, 965–970.

Connine, C. M. (1987). Constraints on interactive processes in auditory word recognition: The role of sentence context. *Journal of Memory and Language, 26*, 527–538.

Connine, C. M., Blasko, D. G., & Hall, M. (1991). Effects of subsequent sentence context in auditory word recognition: Temporal and linguistic constraints. *Journal of Memory and Language, 30*, 234–250.

Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32,* 193–210.

Connine, C. M., & Clifton, C. E. (1987). Interactive use of lexical information in speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 13,* 291–299.

Connine, C. M., Clifton, C. E., & Cutler, A. (1987). Lexical stress effects on phonetic categorization. *Phonetica, 44,* 133–146.

Connine, C. M., Mullenix, J., Shernoff, E., & Yelen, J. (1990). Word familiarity and frequency in visual and auditory word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 16,* 1084–1096.

Connine, C. M., Titone, D., & Wang, J. (1993). Auditory word recognition: Extrinsic and intrinsic effects of word frequency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 19,* 81–94.

Cutler, A. (1980). Errors of stress and intonation. In V. A. Fromkin (Ed.), *Errors in linguistic performance: Slips of the tongue, ear, pen and hand* (pp. 67–80). New York: Academic Press.

Cutler, A. (1986). *Forbear* is a homophone: Lexical prosody does not constraint lexical access. *Language and Speech, 29,* 201–220.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language, 31,* 218–236.

Cutler, A., Butterfield, S., & Williams, J. N. (1987). The perceptual integrity of syllabic onsets. *Journal of Memory and Language, 26,* 406–418.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language, 2,* 133–142.

Cutler, A., & Clifton, C. E. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 183–196). Hillsdale, NJ: Erlbaum.

Cutler, A., Mehler, J., Norris, D. G., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25,* 385–400.

Cutler, A., Mehler, J., Norris, D. G., & Segui, J. (1987). Phoneme identification and the lexicon. *Cognitive Psychology, 19,* 141–177.

Cutler, A., & Norris, D. G. (1979). Monitoring sentence comprehension. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 113–134). Hillsdale, NJ: Erlbaum.

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 113–121.

Cutler, A., & Otake, T. (1994). Mora or phoneme? Further evidence for language-specific listening. *Journal of Memory and Language, 33,* 824–844.

Dell, G. S. (1984). Representation of serial order in speech: Evidence from the repeated phoneme effect in speech errors. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 10,* 222–233.

Dell, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review, 93,* 283–321.

Dell, G. S. (1988). The retrieval of phonological forms in production: Tests of predictions from a connectionist model. *Journal of Memory and Language, 27,* 124–142.

Dell, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language and Cognitive Processes, 5,* 313–349.

Dell, G. S., & Newman, J. E. (1980). Detecting phonemes in fluent speech. *Journal of Verbal Learning and Verbal Behavior, 19,* 609–623.

Dell, G. S., & O'Seaghdha, P. G. (1991). Mediated and convergent lexical priming in language production: A comment on Levelt et al. (1991). *Psychological Review, 98,* 604–614.

Dell, G. S., & O'Seaghdha, P. G. (1992). Stages of lexical access in language production. *Cognition, 42,* 287–314.

Dell, G. S., & Reich, P. (1981). Stages in speech production: An analysis of speech error data. *Journal of Verbal Learning and Verbal Behavior, 20,* 611–629.

Dijkstra, T., Schreuder, R., & Frauenfelder, U. H. (1989). Grapheme context effects on phonemic processing. *Language and Speech, 32,* 89–108.

Dupoux, E., & Mehler, J. (1990). Monitoring the lexicon with normal and compressed speech: Frequency effects and the prelexical code. *Journal of Memory and Language, 29,* 316–335.

Eimas, P. D., Hornstein, S. B. M., & Payton, P. (1990). Attention and the role of dual codes in phoneme monitoring. *Journal of Memory and Language, 29,* 160–180.

Eimas, P. D., & Nygaard, L. C. (1992). Contextual coherence and attention in phoneme monitoring. *Journal of Memory and Language, 31,* 375–395.

Ellis, L., Derbyshire, A. J., & Joseph, M. E. (1971). Perception of electronically gated speech. *Language and Speech, 14,* 229–240.

Elman, J. L. (1990). Representation and structure in connectionist models. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 345–382). Cambridge, MA: MIT Press.

Elman, J. L., & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language, 27,* 143–165.

Elman, J. L., & Zipser, D. (1988). Learning the hidden structure of speech. *Journal of the Acoustical Society of America, 83,* 1615–1626.

Emmorey, K. D. (1989). Auditory morphological priming in the lexicon. *Language and Cognitive Processes, 4,* 73–92.

Fay, D., & Cutler, A. (1977). Malapropisms and the structure of the mental lexicon. *Linguistic Inquiry, 8,* 505–520.

Flores d'Arcais, G., & Schreuder, R. (1987). Semantic activation during object naming. *Psychological Research, 49,* 153–159.

Foss, D. J. (1969). Decision processes during sentence comprehension: Effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behavior, 8,* 457–462.

Foss, D. J. (1982). A discourse on semantic priming. *Cognitive Psychology, 14,* 590–607.

Foss, D. J., & Blank, M. A. (1980). Identifying the speech codes. *Cognitive Psychology, 12,* 1–31.

Foss, D. J., & Gernsbacher, M. A. (1983). Cracking the dual code: Toward a unitary model of phonetic identification. *Journal of Verbal Learning and Verbal Behavior, 22,* 609–632.

Foss, D. J., Harwood, D., & Blank, M. A. (1980). Deciphering decoding decisions: Data and devices. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 165–199). Hillsdale, NJ: Erlbaum.

Foss, D. J., & Ross, J. (1983). Great expectations: Context effects during sentence processing. In G. B. Flores d'Arcais & R. J. Jarvella (Eds.), *The process of language understanding* (pp. 169-191) Chichester: Wiley.

Foss, D. J., & Swinney, D. A. (1973). On the psychological reality of the phoneme: Perception, identification and consciousness. *Journal of Verbal Learning and Verbal Behavior, 12,* 246–257.

Fox, R. A. (1984). Effect of lexical status on phonetic categorization. *Journal of Experimental Psychology: Human Perception and Performance, 10,* 526–540.

Fox, R. A., & Unkefer, J. (1985). The effect of lexical status on the perception of tone. *Journal of Chinese Linguistics, 13,* 69–90.

Frauenfelder, U. H., & Segui, J. (1989). Phoneme monitoring and lexical processing: Evidence for associative context effects. *Memory & Cognition, 17,* 134–140.

Frauenfelder, U. H., Segui, J., & Dijkstra, T. (1990). Lexical effects in phonemic processing: Facilitatory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance, 16,* 77–91.

Fromkin, V. A. (1971). The non-anomalous nature of anomalous utterances. *Language, 47,* 27–52.

Fromkin, V. A. (1973). *Speech errors as linguistic evidence.* The Hague: Mouton.

Fujimura, O., & Lovins, J. B. (1978). Syllables as concatenative phonetic units. In A. Bell & J. B. Hooper (Eds.), *Syllables and segments* (pp. 107–120). Amsterdam: North-Holland.

Ganong, W. F. (1980). Phonetic categorization in auditory word perception. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 110–125.

Garrett, M. F. (1975). The analysis of sentence production. In G. H. Bower (Ed.), *The psychology of learning and motivation* (Vol. 9, pp. 133–177). New York: Academic Press.

Garrett, M. F. (1988). Processes in language production. In F. J. Newmeyer (Ed.), *The Cambridge Linguistic Survey: Vol. 3. Psychological and biological aspects of language* (pp. 69–96) Cambridge, UK: Cambridge University Press.

Glaser, W. R. (1992). Picture naming. *Cognition, 42,* 61–105.

Goldinger, S. D., Luce, P. A., & Pisoni, D. B. (1989). Priming lexical neighbours of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28,* 501–518.

Goldinger, S. D., Luce, P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken word recognition: The roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 18,* 1211–1238.

Goodman, J. C., & Huttenlocher, J. (1988). Do we know how people identify spoken words? *Journal of Memory and Language, 27,* 684–698.

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics, 43,* 137–146.

Gordon, P. C., Eberhardt, J. L., & Rueckl, J. G. (1993). Attentional modulation of the phonetic significance of acoustic cues. *Cognitive Psychology, 25,* 1–42.

Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance,* 21.

Greenspan, S. L. (1986). Semantic flexibility and referential specificity of concrete nouns. *Journal of Memory and Language, 25,* 539–557.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28,* 267–283.

Grosjean, F. (1985). The recognition of words after their acoustic offset: Evidence and implications. *Perception & Psychophysics, 38,* 299–310.

Grosjean, F., & Gee, J. (1987). Prosodic structure and spoken word recognition. *Cognition, 25,* 135–155.

Grosjean, F., & Itzler, J. (1984). Can semantic constraint reduce the role of word frequency during spoken-word recognition? *Bulletin of the Psychonomic Society, 22,* 180–182.

Harley, T. A. (1984). A critique of top-down independent levels of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8,* 191–219.

Healy, A., & Cutting, J. (1976). Units of speech perception: Phoneme and syllable. *Journal of Verbal Learning and Verbal Behavior, 15,* 73–83.

Howes, D. (1957). On the relation between the intelligibility and frequency of occurrence of English words. *Journal of the Acoustical Society of America, 29,* 296–305.

Howes, D. (1966). A word count of spoken English. *Journal of Verbal Learning and Verbal Behavior, 5,* 572–604.

Huttenlocher, J., & Kubicek, L. F. (1983). The source of relatedness effects on naming latency. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 9,* 486–496.

Jarvella, R. J., & Meijers, G. (1983). Recognizing morphemes in spoken words: Some evidence for a stem-organized mental lexicon. In G. B. Flores d'Arcais & R. J. Jarvella (Eds.), *The process of language understanding* (pp. 81–112). Chichester: Wiley.

Jescheniak, J.-D., & Levelt, W. J. M. (1994). Word frequency effects in speech production:

Retrieval of syntactic information and of phonological form. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 824–843.

Katz, L., Boyce, S., Goldstein, L., & Lukatela, G. (1987). Grammatical information effects in auditory word recognition. *Cognition, 25,* 235–263.

Kempen, G., & Huijbers, P. (1983). The lexicalisation process in sentence production and naming: Indirect election of words. *Cognition, 14,* 185–209.

Kempley, S. T., & Morton, J. (1982). The effects of priming with regularly and irregularly related words in auditory word recognition. *British Journal of Psychology, 73,* 441–454.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7,* 279–312.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169–226). Cambridge, MA: MIT Press.

Kohn, S. E., Wingfield, A., Menn, L., Goodglass, H., Gleason, J. B., & Hyde, M. (1987). Lexical retrieval: The tip-of-the-tongue phenomenon. *Applied Psycholinguistics, 8,* 245–266.

Kolinsky, R., Morais, J., & Cluytens, M. (1995). Intermediate representations in spoken word recognition: Evidence from word illusions. *Journal of Memory and Language, 34,* 19–40.

Kubozono, H. (1989). The mora and syllable structure in Japanese: Evidence from speech errors. *Language & Speech, 32,* 249–278.

Lackner, J. R., & Garrett, M. F. (1972). Resolving ambiguity: Effects of biasing context in the unattended ear. *Cognition, 1,* 359–372.

Ladefoged, P., & Broadbent, D. E. (1960). Perception of sequence in auditory events. *Quarterly Journal of Experimental Psychology, 12,* 162-170.

Lahiri, A., & Marslen-Wilson, W. D. (1991). The mental representation of lexical form: A phonological approach to the recognition lexicon. *Cognition, 38,* 245–294.

Levelt, W. J. M. (1989). *Speaking.* Cambridge, MA: MIT Press.

Levelt, W. J. M. (1992). Accessing words in speech production: Stages, processes and representations. *Cognition, 42,* 1–22.

Levelt, W. J. M. (1993). Timing in speech production with special reference to word form encoding. *Annals of the New York Academy of Sciences, 682,* 283–295.

Levelt, W. J. M., & Maassen, B. (1981). Lexical search and order of mention in sentence production. In W. Klein & W. J. M. Levelt (Eds.), *Crossing the boundaries in linguistics: Studies presented to Manfred Bierwisch* (pp. 221–252). Dordrecht: Reidel.

Levelt, W. J. M., Schriefers, H., Vorberg, D., Meyer, A. S., Pechmann, T., & Havinga, J. (1991). The time course of lexical access in speech production: A study of picture naming. *Psychological Review, 98,* 122–142.

Levitt, A. G., & Healy, A. F. (1985). The role of phoneme frequency, similarity and availability in the experimental elicitation of speech errors. *Journal of Memory and Language, 24,* 717–733.

Lucas, M. M. (1987). Frequency effects on the processing of ambiguous words in sentence contexts. *Language & Speech, 30,* 25–46.

Luce, P. A. (1986). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics, 39,* 155–158.

Luce, P. A., Pisoni, D. B., & Goldinger, S. D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 122–147). Cambridge MA: MIT Press.

Lupker, S. J. (1979). The semantic nature of competition in the picture-word interference task. *Canadian Journal of Psychology, 36,* 485–495.

MacKay, D. M. (1972). The structure of words and syllables: Evidence from errors in speech. *Cognitive Psychology, 3,* 210–227.

MacKay, D. M. (1976). On the retrieval and lexical structure of verbs. *Journal of Verbal Learning and Verbal Behavior, 15,* 169–182.

MacKay, D. M. (1987). *The organization of perception and action.* New York: Springer.

Mann, V. A., & Repp, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America, 69,* 548–558.

Marcus, S. M. (1981). ERIS—Context-sensitive coding in speech perception. *Journal of Phonetics, 9,* 197–220.

Marcus, S. M., & Frauenfelder, U. H. (1985). Word recognition—uniqueness or deviation? A theoretical note. *Language and Cognitive Processes, 1,* 163–169.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science, 189,* 226–228.

Marslen-Wilson, W. D. (1980). Speech understanding as a psychological process. In J. D. Simon (Ed.), *Spoken language generation and recognition* (pp. 39–67). Dordrecht: Reidel.

Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X: Control of language processes* (pp. 125–150). Hillsdale, NJ: Erlbaum.

Marslen-Wilson, W. D. (1985). Speech shadowing and speech comprehension. *Speech Communication, 4,* 55–73.

Marslen-Wilson, W. D. (1987). Parallel processing in spoken word recognition. *Cognition, 25,* 71–102.

Marslen-Wilson, W. D. (1990). Activation, competition and frequency in lexical access. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 148–172). Cambridge MA: MIT Press.

Marslen-Wilson, W. D., Brown, C. M., & Tyler, L. K. (1988). Lexical representations in spoken language comprehension. *Language and Cognitive Processes, 3,* 1–16.

Marslen-Wilson, W. D., & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8,* 1–71.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10,* 29–63.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 576–585.

Massaro, D. W. (1987). *Speech perception by ear and eye: A paradigm for psychological inquiry.* Hillsdale NJ: Erlbaum.

Massaro, D. W. (1989). Testing between the TRACE model and the Fuzzy Logical Model of Speech Perception. *Cognitive Psychology, 21,* 398–421.

Massaro, D. W., & Cohen, M. M. (1983). Phonological context in speech perception. *Perception & Psychophysics, 34,* 338–348.

McClelland, J. L., & Elman, J. L. (1986a). Exploiting lawful variability in the speech wave. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 360–385). Hillsdale, NJ: Erlbaum.

McClelland, J. L., & Elman, J. L. (1986b). The TRACE model of speech perception. *Cognitive Psychology, 18,* 1–86.

McNeill, D., & Lindig, K. (1973). The perceptual reality of phonemes, syllables, words, and sentences. *Journal of Verbal Learning and Verbal Behavior, 12,* 431–461.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorisation: Stimulus quality in word-final ambiguity. *Journal of Experimental Psychology: Human Perception and Performance, 17,* 433–443.

McQueen, J. M. (1993). Rhyme decisions to spoken words and nonwords. *Memory & Cognition, 21,* 210–222.

McQueen, J. M., Norris, D. G., & Cutler, A. (1994). Competition in spoken word recognition: Spotting words in other words. *Journal of Experimental Psychology: Learning, Memory and Cognition, 20,* 621–638.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20,* 298–305.

Mehler, J., & Segui, J. (1987). English and French speech processing. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 405–418). Dordrecht: Martinus Nijhoff.

Meyer, A. S. (1990). The time course of phonological encoding in language production: The encoding of successive syllables of a word. *Journal of Memory and Language, 29,* 524–545.

Meyer, A. S., & Schriefers, H. (1991). Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. *Journal of Experimental Psychology: Learning, Memory and Cognition, 17,* 1146–1160.

Milberg, W., Blumstein, S., & Dworetzky, B. (1988). Phonological factors in lexical access: Evidence from an auditory lexical decision task. *Bulletin of the Psychonomic Society, 26,* 305–308.

Miller, J. L. (1981a). Phonetic perception: Evidence for context-dependent and context-independent processing. *Journal of the Acoustical Society of America, 69,* 822–831.

Miller, J. L. (1981b). Some effects of speaking rate on phonetic perception. *Phonetica, 38,* 159–180.

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 369–378.

Miller, J. L., Green, K., & Schermer, T. M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception & Psychophysics, 36,* 329–337.

Miller, J. L., & Liberman, A. L. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25,* 457–465.

Mills, C. B. (1980a). Effects of context on reaction time to phonemes. *Journal of Verbal Learning and Verbal Behavior, 19,* 75–83.

Mills, C. B. (1980b). Effects of the match between listener expectancies and coarticulatory cues on the perception of speech. *Journal of Experimental Psychology: Human Perception and Performance, 6,* 528–535.

Morton, J. (1969). Interaction of information in word perception. *Psychological Review, 76,* 165–178.

Morton, J., & Long, J. (1976). Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behavior, 15,* 43–51.

Morton, J., & Patterson, K. E. (1980). A new attempt at an interpretation, or an attempt at a new interpretation. In M. Coltheart, K. E. Patterson, & J. C. Marshall (Eds.), *Deep dyslexia* (pp. 91–118). London: Routledge & Kegan Paul.

Newman, J. E., & Dell, G. S. (1978). The phonological nature of phoneme monitoring: A critique of some ambiguity studies. *Journal of Verbal Learning and Verbal Behavior, 17,* 359–374.

Nooteboom, S. G. (1981). Lexical retrieval from fragments of spoken words: Beginnings versus endings. *Journal of Phonetics, 9,* 407–424.

Nooteboom, S. G., & van der Vlugt, M. J. (1988). A search for a word-beginning superiority effect. *Journal of the Acoustical Society of America, 84,* 2018–2032.

Norris, D. G. (1982). Autonomous processes in comprehension: A reply to Marslen-Wilson and Tyler. *Cognition, 11,* 97–101.

Norris, D. G. (1986). Word recognition: Context effects without priming. *Cognition, 22,* 93–136.

Norris, D. G. (1990). A dynamic-net model of human speech recognition. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 87–104). Cambridge MA: MIT Press.

Norris, D. G. (1991). Rewiring lexical networks on the fly. *Proceedings of EUROSPEECH 91, Genoa, 1,* 117–120.

Norris, D. G. (1994). SHORTLIST: A connectionist model of continuous speech recognition. *Cognition, 52,* 189–234.

Norris, D. G., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition,* 21.

Oden, G. C., & Spira, J. L. (1983). Influence of context on the activation and selection of ambiguous word senses. *Quarterly Journal of Experimental Psychology, 35,* 51–64.

Oldfield, R. C. (1966). Things, words, and the brain. *Quarterly Journal of Experimental Psychology, 18,* 340–353.

Oldfield, R. C., & Wingfield, A. (1964). The time it takes to name an object. *Nature (London), 202,* 1031–1032.

Oldfield, R. C., & Wingfield, A. (1965). Response latencies in naming objects. *Quarterly Journal of Experimental Psychology, 17,* 273–281.

Onifer, W., & Swinney, D. A. (1981). Accessing lexical ambiguities during sentence comprehension: Effects of frequency-of-meaning and contextual bias. *Journal of Verbal Learning and Verbal Behavior, 17,* 225–236.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32,* 358–378.

Pallier, C., Sebastian-Gallés, N., Felguera, T., Christophe, A., & Mehler, J. (1993). Attentional allocation within the syllabic structure of spoken words. *Journal of Memory and Language, 32,* 373–389.

Pisoni, D. B., & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition, 25,* 21–52.

Pitt, M. A., & Samuel, A. G. (1990). Attentional allocation during speech perception: How fine is the focus? *Journal of Memory and Language, 29,* 611–632.

Pitt, M. A., & Samuel, A. G. (1993). An empirical and meta-analytic evaluation of the phoneme identification task. *Journal of Experimental Psychology: Human Perception and Performance, 19,* 699–725.

Repp, B. H., & Frost, R. (1988). Detectability of words and nonwords in two kinds of noise. *Journal of the Acoustical Society of America, 84,* 1929–1932.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42,* 107–142.

Roelofs, A. (1993). Testing a non-decompositional theory of lemma retrieval in speaking: Retrieval of verbs. *Cognition, 47,* 59–87.

Rubin, D. C. (1975). Within-word structure in the tip-of-the-tongue phenomenon. *Journal of Verbal Learning and Verbal Behavior, 14,* 392–397.

Rubin, P., Turvey, M. T., & van Gelder, P. (1976). Initial phonemes are detected faster in spoken words than in non-words. *Perception & Psychophysics, 19,* 394–398.

Salasoo, A., & Pisoni, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language, 24,* 210–231.

Samuel, A. G. (1981a). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110,* 474–494.

Samuel, A. G. (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance* 7, 1124–1131.

Samuel, A. G. (1987). Lexical uniqueness effects on phonemic restoration. *Journal of Memory and Language, 26,* 36–56.

Samuel, A. G. (1989). Insights from a failure of selective adaptation: Syllable-initial and syllable-final consonants are different. *Perception & Psychophysics, 45,* 485–493.

Samuel, A. G. (1991). A further examination of attentional effects in the phonemic restoration illusion. *Quarterly Journal of Experimental Psychology, 43A,* 679–699.

Samuel, A. G., & Ressler, W. H. (1986). Attention within auditory word perception: Insights from the phoneme restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance, 12,* 70–79.

Savin, H. B. (1963). Word-frequency effect and errors in the perception of speech. *Journal of the Acoustical Society of America, 35,* 200–206.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time course of lexical access in production: Picture-word interference studies. *Journal of Memory and Language, 29,* 86–102.

Schriefers, H., Zwitserlood, P., & Roelofs, A. (1991). The identification of morphologically complex spoken words: Continuous processing or decomposition. *Journal of Memory and Language, 30,* 26–47.

Sebastian-Gallés, N., Dupoux, E., Segui, J., & Mehler, J. (1992). Contrasting syllabic effects in Catalan and Spanish. *Journal of Memory and Language, 31,* 18–32.

Segui, J., & Frauenfelder, U. H. (1986). The effect of lexical constraints upon speech perception. In F. Klix & H. Hagendorf (Eds.), *Human memory and cognitive abilities* (pp. 795–808). Amsterdam: North-Holland.

Segui, J., Frauenfelder, U. H., & Mehler, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology, 72,* 471–477.

Seidenberg, M. S., & Tanenhaus, M. K. (1979). Orthographic effects on rhyme monitoring. *Journal of Experimental Psychology: Human Learning and Memory, 5,* 546–554.

Seidenberg, M. S., Tanenhaus, M. K., Leiman, J. M., & Bienkowski, M. (1982). Automatic access of the meanings of ambiguous words in context: Some limitations of knowledge-based processing. *Cognitive Psychology, 14,* 489–537.

Shallice, T., McLeod, P., & Lewis, K. (1985). Isolating cognitive modules with the dual-task paradigm: Are speech perception and production separate processes? *Quarterly Journal of Experimental Psychology, 37A,* 507–532.

Shattuck-Hufnagel, S. (1979). Speech errors as evidence for a serial-ordering mechanism in sentence production. In W. E. Cooper & E. C. T. Walker (Eds.), *Sentence processing: Psycholinguistic studies presented to Merrill Garrett* (pp. 295–342). Hillsdale, NJ: Erlbaum.

Shattuck-Hufnagel, S. (1983). Sublexical units and suprasegmental structure in speech production planning. In P. F. MacNeilage (Ed.), *The production of speech* (pp. 109–136). New York: Springer-Verlag.

Shattuck-Hufnagel, S. (1986). The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech. *Phonology Yearbook, 3,* 117–149.

Shattuck-Hufnagel, S. (1992). The role of word structure in segmental serial ordering. *Cognition, 42,* 213–259.

Shattuck-Hufnagel, S., & Klatt, D. H. (1979). The limited use of distinctive features and markedness in speech production: Evidence from speech error data. *Journal of Verbal Learning and Verbal Behavior, 18,* 41–55.

Shillcock, R. C. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing* (pp. 24–49). Cambridge MA: MIT Press.

Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language & Speech, 33,* 47–68.

Slowiaczek, L. M., Nusbaum, H. C., & Pisoni, D. B. (1987). Acoustic-phonetic priming in auditory word recognition. *Journal of Experimental Psychology: Learning, Memory and Cognition, 13,* 64–75.

Stemberger, J. P. (1985). An interactive activation model of language production. In A. W. Ellis (Ed.), *Progress in the psychology of language* (Vol. 1, pp. 143–186). London: Erlbaum.

Stemberger, J. P. (1990). Wordshape errors in language production. *Cognition, 35,* 123–157.

Stemberger, J. P. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory and Language, 30,* 161–185.

Stemberger, J. P., Elman, J. L., & Haden, P. (1985). Interference between phonemes during phoneme monitoring: Evidence for an interactive activation model of speech perception. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 475–489.

Stemberger, J. P., & Lewis, M. (1986). Reduplication in Ewe: Morphological accommodation to phonological errors. *Phonology Yearbook, 3,* 151–160.

Stemberger, J. P., & MacWhinney, B. (1986a). Form-oriented inflectional errors in language processing. *Cognitive Psychology, 18,* 329–354.

Stemberger, J. P., & MacWhinney, B. (1986b). Frequency and the lexical storage of regularly inflected forms. *Memory & Cognition, 14,* 17–26.

Sternberg, S., Knoll, R. L., Monsell, S., & Wright, C. E. (1988). Motor programs and hierarchical organization in the control of rapid speech. *Phonetica, 45,* 175–197.

Stevens, K. N. (1986). Models of phonetic recognition II: A feature-based model of speech recognition. In P. Mermelstein (Ed.), *Proceedings of the Montreal Satellite Symposium on Speech Recognition, Twelfth International Congress on Acoustics* (pp. 67–68). Montréal.

Stevens, K. N., Manuel, S. Y., Shattuck-Hufnagel, S., & Liu, S. (1992). Implementation of a model for lexical access based on features. *Proceedings of the Second International Conference on Spoken Language Processing, Banff, 1,* 499–502.

Swinney, D. A. (1979). Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior, 18,* 645–659.

Swinney, D. A., Onifer, W., Prather, P., & Hirschkowitz, M. (1979). Semantic facilitation across sensory modalities in the processing of individual words and sentences. *Memory & Cognition, 7,* 159–165.

Swinney, D. A., & Prather, P. (1980). Phonemic identification in a phoneme monitoring experiment: The variable role of uncertainty about vowel contexts. *Perception & Psychophysics, 27,* 104–110.

Tabossi, P. (1988a). Accessing lexical ambiguity in different types of sentential contexts. *Journal of Memory and Language, 27,* 324–340.

Tabossi, P. (1988b). Effects of context on the immediate interpretation of unambiguous nouns. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 14,* 153–162.

Tabossi, P., Burani, C., & Scott, D. R. (1995). Word identification in fluent speech. *Journal of Memory and Language,* 34.

Tabossi, P., Colombo, L., & Job, R. (1987). Accessing lexical ambiguity: Effects of context and dominance. *Psychological Research, 49,* 161–167.

Taft, M. (1986). Lexical access codes in visual and auditory word recognition. *Language and Cognitive Processes, 1,* 297–308.

Taft, M. (1988). A morphological-decomposition model of lexical representation. *Linguistics, 26,* 657–667.

Taft, M., & Hambly, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition, 22,* 259–282.

Taft, M., Hambly, G., & Kinoshita, S. (1986). Visual and auditory recognition of prefixed words. *Quarterly Journal of Experimental Psychology, 38A,* 357–366.

Tanenhaus, M. K., & Donnenwerth-Nolan, S. (1984). Syntactic context and lexical access. *Quarterly Journal of Experimental Psychology, 36A,* 649–661.

Tanenhaus, M. K., Leiman, J. M., & Seidenberg, M. S. (1979). Evidence for multiple stages in the processing of ambiguous words in syntactic contexts. *Journal of Verbal Learning and Verbal Behavior, 18,* 427–440.

Treiman, R. (1983). The structure of spoken syllables: Evidence from novel word games. *Cognition, 15,* 49–74.

Treiman, R. (1986). The division between onsets and rimes in English syllables. *Journal of Memory and Language, 25,* 476–491.

Tyler, L. K. (1984). The structure of the initial cohort: Evidence from gating. *Perception & Psychophysics, 36,* 417–427.

Tyler, L. K., Marslen-Wilson, W. D., Rentoul, J., & Hanney, P. (1988). Continuous and discontinuous access in spoken word recognition: The role of derivational prefixes. *Journal of Memory and Language, 27,* 368–381.

Tyler, L. K., & Wessels, J. (1983). Quantifying contextual contributions to word recognition processes. *Perception & Psychophysics, 34,* 409–420.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance,* 21.

Warren, R. M. (1970). Perceptual restoration of missing speech sounds. *Science, 167,* 392–393.

Whalen, D. H. (1984). Subcategorical mismatches slow phonetic judgments. *Perception & Psychophysics, 35,* 49–64.

Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics, 50,* 351–360.

Whitney, P., McKay, T., Kellas, G., & Emerson, W. A. (1985). Semantic activation of noun concepts in context. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 11,* 126–135.

Williams, J. N. (1988). Constraints upon semantic activation during sentence comprehension. *Language and Cognitive Processes, 3,* 165–206.

Wingfield, A. (1968). Effects of frequency on identification and naming of objects. *American Journal of Psychology, 81,* 226–234.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition, 32,* 25–64.

Zwitserlood, P., Schriefers, H., Lahiri, A., & van Donselaar, W. (1993). The role of syllables in the perception of spoken Dutch. *Journal of Experimental Psychology: Learning, Memory and Cognition, 19,* 260–271.