Lexical Stress and Lexical Access: Effects in Read and Spontaneous Speech

Janice Margaret McAllister

Ph.D.

University of Edinburgh

1989

Declaration

I declare that I am solely responsible

for the composition of this thesis

and that the work reported herein

was conducted by me.

Abstract

This thesis examines three issues which are of importance in the study of auditory word recognition: the phonological unit which is used to access representations in the mental lexicon; the extent to which hearers can rely on words being identified before their acoustic offsets; and the role of context in auditory word recognition. Three hypotheses which are based on the predictions of the Cohort Model (Marslen-Wilson and Tyler 1980) are tested experimentally using the gating paradigm. First, the phonological access hypothesis claims that word onsets, rather than any other part of the word, are used to access representations in the mental lexicon. An alternative candidate which has been proposed as the initiator of lexical access is the stressed syllable. Second, the early recognition hypothesis states that polysyllabic words, and the majority of words heard in context, will be recognised before their acoustic offsets. Finally, the context-free hypothesis predicts that during the initial stages of the processing of words, no effects of context will be discernible.

Experiment 1 tests all three predictions by manipulating aspects of carefully articulated, read speech. First, examination of the gating responses from three context conditions offers no support for the context-free hypothesis. Second, the high number of words which are identified before their acoustic offsets is consistent with the early recognition hypothesis. Finally, the phonological access hypothesis is tested by manipulation of the stress patterns of stimuli. The dependent variables which are examined relate to the processes of lexical access and lexical retrieval; stress differences are found on access measures but not on those relating to retrieval. When the experiment is replicated with a group of subjects whose level of literacy is lower than that of the undergraduates who took part in the original experiment, differences are found in measures relating to contextual processing.

Experiment 2 continues to examine the phonological access hypothesis, by manipulating speech style (read versus conversational) as well as stress pattern. Gated

words, excised from the speech of six speakers, are presented in isolation. Words excised from read speech and words stressed on the first syllable elicit a greater number of responses which match the stimuli than conversational tokens and words with unstressed initial syllables. Intelligibility differences among the four conditions are also reported.

Experiment 3 aims to investigate the processing of read and spontaneous tokens heard in context, while maintaining the manipulation of stress pattern. A subset of the words from Experiment 2 are presented in their original sentence contexts: the test words themselves, plus up to three subsequent words, are gated. Although the presence of preceding context generally enhances intelligibility, some words remain unrecognised by the end of the third subsequent word. An interaction between stress and speech style may be explained in terms of the unintelligibility of the preceding context.

Several issues arising from Experiments 1, 2 and 3 are considered further. The characteristics of words which fail to be recognised before their offsets are examined using the statistical technique of regression; the contributions of phonetic and phonological aspects of stressed syllables are assessed; and a further experiment is reported which explores top-down processing in spontaneous speech, and which offers support for the interpretation of the results of Experiment 3 offered earlier.

Acknowledgements

Over the years which have elapsed since I began work on this thesis, I have drawn on the help of many people, and I am glad to be able to take this opportunity to thank them formally.

The first group of people whom I must mention are the members of my supervisory committee. First and foremost I acknowledge my debt to my supervisor Ellen Gurman Bard who, luckily for me, possesses wisdom, patience and a sense of humour in roughly equal quantities. Her incisive comments, though not always appreciated at the time, have greatly improved the quality of this thesis. Steve Isard joined in the supervision of this work at a relatively late stage, but the duration of his involvement is in inverse proportion to the amount of helpful advice and encouragement he has provided. I would also like to thank the other members of my supervisory panel, Ron Asher and Jim Hurford.

For technical help of various sorts, I am grateful to Norman Dryden, Irene MacLeod and Stewart Smith.

Many of my colleagues in the Department of Linguistics and at the Centre for Speech Technology Research have been the source of helpful discussion and moral support. In particular, I would like to thank Alan Kemp, John Laver, Jim Miller, Andy Lowe and Richard Shillcock.

The experiments reported here would not have been possible without the cooperation of numerous speakers and subjects who made their time and tolerance available. I am also grateful to the Scottish Education Department for providing the the grant which funded the initial stages of this work.

Any enterprise of this kind has repercussions for the unfortunate family of the author. My son Martin has borne the consequences of my efforts with characteristic good grace for more years than I care to remember. But in keeping with time-honoured tradition, I save my biggest debt till last. Without the support, help and advice of my husband and colleague Mike McAllister, I would never have reached the stage of submitting this thesis. In dedicating it to him I can only hope to repay a tiny part of that debt, but I do so with deepest gratitude.

Table of Contents

CHAPTER ONE

Introduction

The past decade has seen a growth in interest within the field of psycholinguistics in the nature of the mechanisms hearers use to recognise words in continuous speech. This study is concerned with several aspects of the processing of spoken words.

In modelling the word recognition process, researchers customarily recognise at least two distinct processing operations (see, for example, Cutler 1986; Marslen-Wilson 1987). During lexical access, a set of word candidates is assembled. During lexical retrieval, a single word candidate emerges from this set and is recognised. A further dichotomy which is usually recognised in accounts of word recognition concerns two knowledge sources which are termed bottom-up and top-down information. Bottom-up processing is concerned with deriving high-level representations from low-level information: in the case of auditory word recognition, the acoustic signal is processed to yield lexical hypotheses. Top-down processing is concerned with the influence of contextual constraints such as syntax, semantics, morphology and the lexical structure of the language. Models of auditory word recognition must make predictions about the manner in which these two knowledge sources are utilised, and about the relationship between them and the two processing operations identified above.

The question of the nature of bottom-up processing for lexical access has been the focus of much debate. In particular, researchers have sought to identify the phonological unit which is used to access the mental lexicon. One candidate which has been proposed is the word-initial syllable. A common objection to models which advocate initial syllables as a means of accessing the lexicon (for example, those proposed by Marslen-Wilson and Welsh 1978, Marslen-Wilson and Tyler 1980, Cole and Jakimik 1980, Tyler 1984, and Marslen-Wilson 1987) is that they require the beginnings of words to be identified pre-lexically: yet there is a large amount of literature suggesting that although speakers *can* mark word boundary cues by acoustic-phonetic events of various kinds, in connected speech they rarely do so (Brown 1977; Gimson 1980). An alternative means of identifying word boundaries might be the efficient recognition of the words preceding the boundary, so that, as Cole and Jakimik (1980) have suggested, a hearer knows where a word begins because the preceding word has already been recognised. However, Grosjean (1985) and Bard et al. (1988) have shown that this condition is frequently not met in connected speech. Indeed, Bard et al. have demonstrated that in conversational speech a word may not be recognised until later words have been identified. Thus it seems likely that hearers will often be unable to determine which portions of the acoustic signal correspond to a word onset prior to recognising the word. In addition to this segmentation problem, hearers may be faced with further difficulties even if the location of word onsets is known; a single lexical item may be realised in a variety of ways even by the same speaker, so that a simple pattern-matching approach to lexical processing is infeasible.

By contrast, lexically stressed syllables ought to be less susceptible to the problems which beset word onsets. First, if the word containing the lexically stressed syllable is assigned sentence accent, the syllable may be marked by increased amplitude and duration and by fundamental frequency movement (see, for example, Cutler and Ladd 1983), characteristics which might facilitate their processing in the speech signal. Second, syllables bearing lexical stress are relatively invariant, in the sense that they are far less likely than other parts of words to be affected by phonological processes (Brown 1977). A further advantage which lexically stressed syllables possess is that they are more informative in the sense of more efficiently partitioning the lexicon, by virtue of the fact that they permit a larger range of vowels than unstressed syllables (Altmann and Carter 1988).

Clearly the experimental results just outlined relate to a number of different aspects of the stress phenomenon. On the one hand, the abstract property of lexical stress is correlated with characteristics of spoken words which might lead to their being more informative, in the sense of more efficiently partitioning the lexicon, and more reliable, in the sense of being less likely to undergo phonological modification. On the other hand, assignment of sentence stress to a lexically stressed syllable may also lead to its being louder and longer than other syllables and marked by fundamental frequency movement, which might contribute to its being identified and processed relatively easily. Factors such as these have no doubt led to the suggestion that stressed syllables (variously defined) are used to initiate lexical access (Cutler 1976; Grosjean and Gee 1987) or at least have a special role in speech processing (Brown 1977). Although many experimental results have suggested that stressed syllables are in some sense easier to process than unstressed syllables (e.g. Shields, McHugh and Martin 1973; Cutler 1976), there is little direct evidence in favour of the hypothesis that stressed syllables have a special role in lexical access.

The nature of top-down processing has also been the focus of much experimental attention. Theories of word recognition may be characterised according to the extent to which contextual information is used in the identification of words: parallel models allow top-down information to influence the identification process, while serial autonomous models do not. That is, within a serial autonomous model, such as that proposed by Forster (1976, 1979, 1981), words are recognised purely on the basis of their acoustic form. However, several studies (see Marslen-Wilson (1987) for a summary) have demonstrated that words heard in context are frequently recognised before sufficient acoustic-phonetic information has accumulated to distinguish them uniquely from the remainder of the lexicon. Yet even if the existence of the early recognition phenomenon excludes the serial autonomous class of models from consideration (and this is still a matter of some disagreement), there is still room for debate over the issue of the use of context in word recognition, given the range of logically permissible parallel models. One possibility is that top-down processing is permitted during both lexical access and lexical retrieval (Morton 1969). An alternative view, which is still consistent with the notion of parallel processing, is that access itself may be driven by bottom-up considerations alone, but that contextual factors may affect retrieval (Marslen-Wilson and Tyler 1980).

The bottom-up and top-down aspects of lexical processing are clearly open to a number of possible treatments in terms of models of word recognition, and this diversity is reflected in the wide range of models which has been proposed. Of these, the theory which has made the most detailed and testable claims about *both* top-down and bottom-up aspects of processing is the Cohort Model. According to the original formulation of this theory, (Marslen-Wilson and Welsh 1978; Marslen-Wilson and Tyler 1980; Tyler 1984), initial syllables are used to access the mental lexicon, and contextual information is not taken into account until the retrieval stage. A corollary of the former feature is that the majority of words must be recognised by their acoustic offsets, in order to indicate the position of the onsets of following words.

The Cohort Model has recently been modified (Marslen-Wilson 1987) so that instead of requiring a perfect match between input and stored representation, the processor compares descriptions of input and target framed in featural terms. A goodness-of-fit metric is then invoked to establish the identity of the intended word. The role of context has also been modified, so that instead of functioning to eliminate candidates from the cohort, it provides a framework against which the senses associated with candidate words can be assessed. Finally, a mechanism for modelling word frequency effects has also been incorporated.

The aim of this thesis is to test three related predictions of the Cohort model, which concern both top-down and bottom-up aspects of the processing of words. Since the revisions to the model which were described in the previous paragraph took place after the main experiments of this study were conducted, the hypotheses outlined in this chapter are framed in terms of the original version of the theory. However, the implications of the experimental results for the new model will be discussed.

The first claim of the Cohort model which will be considered here is that of the unit which is used to access the mental lexicon, which according to the model is the initial portion (e.g. syllable) of the input. This claim will be evaluated with respect to the alternative proposal for the unit of phonological access, namely the stressed syllable. The Cohort Model's claim that word onsets are used to access the lexicon will be termed the *phonological access hypothesis*. While much existing evidence has favoured word onsets in this role, the results relating to the processing of stressed syllables is equivocal. This may be because the experiments in question have used carefully articulated, read speech. In spontaneous speech, the value of stressed syllables as 'islands of reliability' is likely to be greater.

The second prediction of the Cohort model which this study will examine concerns the eventual recognition of words in context. Marslen-Wilson (1987:75) claims that 'a large proportion of words are selected [i.e. recognised] early'. This *early recognition hypothesis* is in some respects as fundamental to the Cohort model as the phonological access hypothesis just discussed, since the knowledge that one is dealing with the initial portion of a word (and should therefore activate a word initial cohort) is often dependent on the knowledge that the preceding word has ended (Cole and Jakimik 1980). Although the early recognition phenomenon is well established in the psycholinguistic literature, it should be noted that the experiments which support it generally use materials which are very carefully articulated; it is not clear to what extent early recognition is a factor in the processing of, for example, conversational speech. The phenomenon of late recognition, however, is less well understood.

The final issue which will be examined is whether the accessing of the word-initial cohort is affected by the presence of contextual information. The outcome of this investigation is crucial to the Cohort model since the insistence on a bottom-up autonomous phase is one of the features which distinguishes the theory from others which also employ parallel processing. This prediction of the model will be termed the *context-free hypothesis.*

This thesis will be structured as follows. In Chapter 2, the existing literature relating to the word recognition process is discussed. The aims of this survey are two-fold. First, the nature of the input -- normal continuous speech -- is considered with a view to determining the factors which are likely to pose problems for the word recognition process. Second, the experimental work concerning the human word recognition mechanism is examined, in order to establish the empirical underpinnings of the existing word recognition models, which are also considered. The chapter concludes with a detailed presentation of the hypotheses under investigation.

Chapter Three is concerned with the identification of a methodology appropriate to the study being undertaken. The range of available experimental tools is examined and the selection of a single technique, gating (Grosjean 1980), is justified.

The experimental work of this thesis is reported in Chapters 4, 5 and 6. Experiment 1, which is described in Chapter 4, sets out to test all three predictions by manipulating aspects of the carefully articulated, read speech which is commonly used in psycholinguistic experiments. First, examination of the gating responses from three context conditions offers no support for the context-free hypothesis. Second, the high number of successful identifications -- whether words are heard in isolation or in context -- is consistent with the early recognition hypothesis. Finally, the phonological access hypothesis is tested by manipulation of the stress patterns of the stimuli. The dependent variables which are examined relate to both lexical access and lexical retrieval; stress differences are found on the lexical access measures but not on those relating to retrieval.

Chapter 5 describes two experiments which investigate the processing of read and spontaneous speech. Experiment 2 continues to examine the phonological access hypothesis, by manipulating speech style (read versus conversational) as well as stress pattern. Gated words, excised from the speech of six speakers, are presented in isolation. Words excised from read speech and words stressed on the first syllable elicit a greater number of responses which match the stimuli than conversational tokens and words with unstressed initial syllables. These findings pose problems for models claiming that word onsets are used to access the mental lexicon.

Experiment 3 aims to investigate the processing of read and spontaneous tokens heard in context, while maintaining the manipulation of stress pattern. A subset of the words from Experiment 2 are presented in their original sentence contexts; the test words themselves, plus up to three subsequent words, are gated. Although the presence of preceding context generally enhances intelligibility, some words remain unrecognised by the end of the third subsequent word. Once again, the results are inconsistent with the claims of the 1980 version of Cohort. An interaction between the stress and speech style variables may be explained in terms of the acoustic-phonetic ambiguity of the preceding context.

In Chapter 6, several issues arising from Experiments 1, 2 and 3 are considered further. The characteristics of words which fail to be recognised before their offsets are examined using the statistical technique of regression; the contributions of phonetic and phonological aspects of stressed syllables are assessed; and a further experiment is described which explores top-down processing in spontaneous speech, and which offers support for the interpretation of the results of Experiment 3 discussed in Chapter 5.

Several of the experimental findings presented in Chapters 4 and 5 pose problems for the Cohort model -- and indeed for other models of word recognition. Chapter 7 summarises the results of the research, and discusses ways in which the current models of word recognition must be modified to account for them.

CHAPTER 2

Review of the Literature

1. Introduction

The aim of this chapter is to provide the background against which this study is set. Over the past few decades, a copious amount of research into auditory word recognition has been conducted. To attempt to present all of this evidence in this chapter would be both difficult and counter-productive, the latter because it would tend to obscure the very points which are relevant to the thesis. The approach which has been adopted in surveying the literature on word recognition is therefore selective; the evidence which has been included relates to a small number of well-defined aspects of the recognition of words, namely the use of context, the extent to which processing proceeds in a word-by-word fashion, and the manner in which the acoustic-phonetic input is processed to yield a set of word candidates.

The process of word recognition is customarily divided into two subsidiary tasks. The first, frequently termed *lexical access* (e.g. Cutler 1986; Marslen-Wilson 1987; but Tyler (1984) terms this *cohort initiation*), consists of selecting a set of word candidates in the lexicon which are then subject to further processing. The second, *lexical retrieval* (Cutler 1986; or *selection*, according to Marslen-Wilson (1987); or *cohort resolution*, according to Tyler (1984)), involves the emergence of a single candidate as the recognised word (1).

A second dichotomy in the word recognition literature is that between knowledge sources. One knowledge source is clearly the acoustic-phonetic representation of the word being processed: it is generally assumed that a mapping process between the acoustic-phonetic input and stored representation is one (and in some models, *the*) method by which access is gained to an initial set of word candidates in the lexicon. A second source of knowledge is contextual in nature (2); the degree to which contextual information is exploited during the word recognition process, and the precise locus of contextual effects, are major defining characteristics of models of auditory word recognition.

This chapter is organised as follows. Section 2 examines the nature of the input to which the hearer is exposed -- continuous speech -- and concludes that hearers can rely neither on explicit cues to the locations of word boundaries nor on invariant surface manifestations of words. Section 3 is concerned with perceptual evidence relating to word recognition. First, it examines the consequences of the lack of segmentation cues and the variability inherent in connected speech, and shows that these problems complicate the task of word recognition for the hearer. Second, it considers the extent to which contextual information contributes to efficient lexical processing: while many writers have claimed that such cues affect the word recognition process, there is disagreement about the stage at which they are used. Finally, in Section 4, some existing models of word recognition are described, as is the extent to which they meet -- or fail to meet -- the constraints introduced by the effects outlined in sections 2 and 3.

2. Linguistic considerations

In this section, we will examine the nature of the input to listeners -- that is, fluent speech -- with a view to determining what particular problems they will face in attempting to recognise words. As the studies described will demonstrate, two aspects of continuous speech are likely to cause difficulties for hearers: first, the lack of consistent cues to the locations of word boundaries; and second, the absence of an invariant relationship between the surface forms of words and their assumed underlying representation. However, some parts of words (most notably, stressed syllables) are relatively resistant to variability.

2.1. The segmentation problem

The task of word recognition would be greatly facilitated if it were possible to detect word boundary cues which could be readily identified in the acoustic signal. Yet it has long been observed that the onsets of words are not unambiguously marked in this way:

> The word is not primarily a phonetic unit; we do not, by pauses or other phonetic features, mark off those segments of our speech which could be spoken alone (Bloomfield 1933:181)

The situation may not be quite so grim for the hearer as Bloomfield leads us to believe. Several writers (e.g. Brown, 1977; Gimson, 1980) have described word boundary cues which may be present in the formal style of continuous speech. For example, Brown lists features which may signal word onsets, such as aspirated voiced stops in stressed syllables, strong frication in fricatives and a 'light' quality to /l/ sounds. Similarly, word offsets may be characterised by glottalisation of voiceless stops. Another feature associated with word boundaries is consonant gemination (for example, in the dyad *talked to*, when carefully pronounced); the juxtaposition of two identical consonants leads to a lengthening of the stop closure which is not observed word-internally except at compound boundaries.

Lehiste (1960) conducted an analysis of the junctural cues which speakers might use to differentiate pairs such as *a nice man* versus *an ice man* and *I scream* versus *ice cream*. She identified features such as lengthening, amplitude changes, glottal stop insertion and devoicing. Nakatani and Dukes (1977) likewise examined the acoustic-phonetic characteristics of pairs of utterances such as *no notion* / *known ocean* and identified a number of cues which may mark word onsets, such as the insertion of a glottal stop or the use of creaky voice when a word begins with a stressed vowel, allophonic variation for consonants such as /l/ and /r/, and word duration. By editing and splicing the original versions of the utterances and playing hybrid versions to hearers, the authors were able to isolate the effects of individual cues.

These studies may lead us to conclude that a number of acoustic cues to the presence and location of word boundaries are available to the hearer. The artificiality of the conditions under which the materials were produced should not, however, be underestimated. When speakers produce utterances in a list containing other items which are clearly meant to contrast in terms of boundary placement (even when large numbers of filler items are also included) it is hardly surprising if juncture cues are elicited. While it is of interest that the speaker has the option of employing such devices to clarify the intended interpretation of a phoneme sequence, the issue in this discussion is to what extent speakers *habitually* provide such information for hearers. Brown's verdict is that these cues cannot be relied upon in any but the most formal speech style:

> In informal speech these markers are very frequently obscured to some extent. (Brown 1977:72)

Gimson (1980:296) makes a similar point:

> Junctural oppositions are, in fact, frequently neutralised in connected speech or have such slight phonetic value as to be difficult for a listener to perceive.

Vaissiere (1985:207) relates the loss of segmentation cues to an increase in speech rate:

> Acoustic cues for ... word ... boundaries become more and more difficult to detect as the rate of speaking increases.

Lamel and Zue (1984) have explored an alternative approach to the identification of word boundaries. They note that, of approximately 7,000 consonant sequences which could theoretically be observed at word boundaries, only 20% could also occur word-medially. They reason that, if the speech processor encounters such a sequence, it can be confident that a word boundary lies within it. Furthermore, they point out that, in 80% of the consonant sequences which must contain a word boundary, there is only one possible position for it. For example, the sequence /mgl/ (as in *same glass*) must contain a word boundary after the /m/. The opinion of Lamel and Zue seems to be that hearers ought to be able to deduce the locations of boundaries in a high percentage of cases. Unfortunately, their reasoning is flawed because they fail to take into account the processes of assimilation and deletion which will be discussed in the next section. Their analyses are all based on idealised transcriptions of words, and not on the forms which actually occur. Thus, in many cases, the sequence of consonants which they describe would actually be reduced by consonant deletion or modified by assimilation and would no longer constitute an unambiguous word boundary sequence.

In summary, then, the linguistic evidence relating to the availability to the hearer of explicit and unambiguous word boundary cues does not seem very promising. Such cues are very limited in number, and writers are in agreement that the likelihood of their being present diminishes with increasing informality of style and/or speech rate.

2.2. The variability problem

Even if word boundaries could be unambiguously determined (and the evidence cited above indicates that this is rarely possible in informal speech) hearers would still have to confront problems arising from the the variable nature of the speech signal. That is, a given word type may give rise to an enormous number of tokens which bear only an indirect relationship to each other and to the idealised representation of the word type. Yet it is often implicitly assumed by linguists and psycholinguists alike that there is a stable, invariant relationship between related word tokens (for discussion of this point, see Linell 1982).

There are at least three sources of this variability: inter-speaker differences, intra-speaker differences and phonological context. Nolan (1983) has catalogued the kinds of physiological differences between speakers which may result in variability in the surface forms of utterances. For example, the length of the vocal tract is a major determinant of the acoustic characteristics of a vowel, with shorter vocal tracts resulting in higher Hertz values for formants. Peterson and Barney (1952) used first and second formant frequencies to plot vowels uttered by different speakers; when materials and dialect were held constant, there was still a considerable degree of overlap among vowels. Other physiological factors which influence speakers' articulation are the characteristics of the vocal folds and the shape of the nasal tract.

As well as interspeaker differences introduced by physiological factors which are not, presumably, under the speaker's control, a range of effects may be observed which result from habitual settings of the vocal tract which speakers adopt more or less consciously. Laver (1980) identifies a variety of supralaryngeal and phonatory settings which are responsible for differences in perceived voice quality in speakers, and which must contribute substantially to the variability between individual word tokens.

Although physiological differences between speakers account for a large degree of variability between word tokens, the relationship between word tokens uttered by a single speaker is scarcely more stable. Emotional factors such as anger, fear and sorrow may affect the pronunciation of a given word by a given speaker (Williams and Stevens 1972).

A further kind of variability arises from the style of speech adopted by the speaker. Shockey (1983) examined the factors affecting the pronunciations of words by first recording speakers in a conversational setting, then, when she had transcribed the conversation and selected a subset of utterances, having the same speakers record carefully articulated versions of the selected materials. She found marked differences in the realisations of words in the two speech styles, attributable to a more frequent application, in casual speech, of a variety of phonological rules. Although the speech rate variable is implicated in the application of these rules (Dalby 1986) Shockey concludes that the style factor is independent of speech rate.

While the types of inter- and intra- speaker differences just discussed are difficult to account for without reference to extra-linguistic factors, a further kind of variability is motivated -- in part, at least -- by features of the phonological context itself.

Variability arising from phonological context has been widely documented in the literature. For example, Gimson (1980) and Brown (1977) have discussed phonological modifications observed in RP English. The types of modification which speakers introduce are many and varied. Processes affecting consonants include glottalisation, palatalisation, flapping, weaking, shortening, gemination and deletion, while vowels may be altered in terms of duration, amplitude and spectral characteristics. The consequence of this degree of variability at the segmental level is an enormous range of potential pronunciations for individual words. For example, Stampe (1979) lists nine possible pronunciations for the word *divinity* in American English, and claims that this does not exhaust the range of possibilities. Harrington et al. (1988) have reported similar variation for RP English.

If, as several authors imply, the application of the word boundary phonology rules discussed above is optional (that is, entirely dependent on the whim of the speaker), the consequences for the hearer are potentially serious, since it will be impossible to predict the occurrence of a rule (and hence, to 'undo' it). However, more recent work has indicated that the process is far more constrained than was originally thought. Cooper and Paccia-Cooper (1980) examined the relationship between lexical, prosodic and syntactic factors and the surface realisations of spoken words. They discovered that processes such as palatalisation and alveolar flapping were more likely to be suspended before major syntactic boundaries, the site of verb deletions, words bearing emphatic stress and low frequency words (3). Noting that these factors did not account for all the observed variability in the data, Cooper et al. (1983) examined the effects of speech rate on the occurrence of the palatalisation process. They found that palatalisation across word boundaries was more frequent among characteristically fast speakers and at fast rates of speech than elsewhere.

Cooper and Paccia-Cooper found that syllable duration was also systematically affected by high-level factors of the kind just described. Further research into the relationship between sentence and discourse level factors and duration is reported by Fowler and Housum (1987). They examined the duration of word tokens which constituted either the first or the second mention of the word in a particular text, and found that the first token was relatively longer than the second. In addition, there was a significant correlation between word length in milliseconds and the likelihood of attenuation, and words closely related to the topic of the sentence shortened less. The distance between the first and second token was irrelevant, however.

Writers on phonological modifications seem to agree that the frequency of such modifications increases with the adoption of progressively more casual styles of speech. Conversational speech, at one end of the spectrum, differs from read speech in a number of ways. Many of these differences are suprasegmental in nature. Johns-Lewis (1986) has found that speakers produce spontaneous speech at a slower rate than read speech; this is no doubt attributable to the longer pauses and hesitations in the former (Crystal and Davy 1969), rather than the length of the word tokens themselves. A further prosodic characteristic is that the fundamental frequency range is narrower (Johns-Lewis 1986). When subjects hear utterances drawn from read and spontaneous samples, they can distinguish the two styles, but they can do so less accurately when basing their judgments on transcripts of the materials (Johns-Lewis 1987).

2.2.1. Stressed syllables and phonological modification

Interestingly, certain parts of words are less vulnerable than others to the phonological processes just discussed. It is often said that lexically stressed syllables are relatively immune to phonological modification. The issue of stress has interested phonologists for decades (Fry,1958; Kingdon 1958; Chomsky and Halle 1968; Hyman 1977; Fudge 1984; Mohanan 1986). At an abstract level, English polysyllabic words have at least one syllable which is marked as bearing the suprasegmental feature of lexical stress, which may be signalled by greater amplitude and duration and by an excursion of the fundamental frequency contour (Adams and Munro 1977; Ladefoged

1982). It also correlates with the occurrence of full vowels, since the vowel reduction rules of English are blocked by the presence of lexical stress (Fudge 1984).

Phonological theories of lexical stress assignment, and related psycholinguistic studies, have given rise to a confusing array of terminology. The term 'stress' refers to an abstract property assigned to particular syllables when lexical stress is at issue (syllables may be assigned primary or secondary stress); but it is also used to indicate portions of multi-word utterances which receive particular emphasis, in the case of sentence stress or accent. Another parameter which is relevant to the analysis of stress is syllable weight, which is defined in terms of the segmental composition of syllables (the number of consonants following the vowel, and the length of the vowel itself); syllables may be 'weak' or 'strong' (4). The presence of stress is one of the major determinants of vowel quality; if syllables are stressed, they will contain full vowels, but if not, it is likely that they will contain reduced vowels (schwa and /ɪ/).

Lexical stress, which results in a particular stress pattern when words are produced in isolation, must be distinguished from sentence stress (or sentence accent) which is a characteristic of multi-word utterances. Although a syllable may bear lexical stress, when it is uttered in connected speech it may fail to be realised as stressed (in terms of the acoustic factors of amplitude, duration and F_0 modulation), although the unreduced vowel quality will tend to be preserved to some degree. While the location of lexical stress in English is determined by the morphological and phonological structure of the word, sentence accent placement depends on syntactic, semantic and pragmatic factors. Languages like English and Dutch, in which accented syllables occur at roughly regular intervals, are termed *stress-timed languages* (Abercrombie 1967), and are contrasted with *syllable-timed languages* such as French, in which syllables occur at roughly equal intervals. Words spoken in sentence contexts whose stressed syllables are realised with the physical features of increased amplitude and duration and marked fundamental frequency movement are said to be accented. The issue of determining the location of sentence accent is widely debated (5).

Brown (1977:55) notes that the sorts of segmental phonological modifications which she discusses are usually considered unacceptable if they occur in stressed syllables, while conversely, they are frequently found in unstressed syllables. Brown repeatedly qualifies the discussion of assimilatory processes by pointing out that stressed syllables are relatively immune, e.g.

> In a stressed syllable the initial consonant(s) and the vowel will be very clearly enunciated whereas in unstressed syllables the consonants may be very weakly enunciated and the vowel very obscure. (1977:46)
>
> 'voiced' consonants which occur in unstressed syllables [become] 'voiceless' when, as a result of the elision of /ə/, they occur next to a 'voiceless' consonant. (1977:69)
>
> With the single exception of /ˈpaps/ [the word *perhaps*] no stressed syllable [in the corpus] is affected by elision – what tends to happen is that a series of unstressed syllables are run together. (1977:70)
>
> 'weakening' is very common. The most frequent examples are of 'weakening' of /k/ and /t/. The general requirement seems to be simply that the consonant should not be initial in a stressed syllable. (1977:74)
>
> vowels in unstressed syllables tend to simplify. (1977:75)
>
> /ju/ and /u/ are fairly stable in stressed syllables ... the /i/-ending set of vowels is relatively stable in stressed syllables. (1977:78)

Other writers have noted the relative stability of stressed syllables. For example, Peterson and Lehiste (1960) demonstrated that stressed vowels shorten less than unstressed vowels. Observations such as these lead Brown to speculate (p. 60) that

> in listening to spoken English the native speaker concentrates on the stressed syllables.

Whether Brown's hypothesis is correct will be discussed in Section 3, when perceptual studies will be examined.

However, it is not only because of stressed syllables' relative immunity to phonological reduction that writers have suggested that these parts of words play a crucial role in speech recognition. It has been shown, for example, that stressed syllables are louder and longer than others (Lehiste 1967; Adams and Munro 1977; Umeda 1977; Ladefoged 1982), attributes which we might expect would enhance listeners' ability to process them.

Huttenlocher and Zue (1983) have suggested another reason for considering stressed syllables as important elements in the word recognition process. They used the broad phonetic class analysis which had been proposed by Shipman and Zue (1982), who had advocated conducting a lexical search using manner-based descriptions of the segments of words and had found that this approach could result in an efficient lexical search procedure in automatic, large vocabulary, isolated word recognition. The strategy proposed by Shipman and Zue involved analysing the input word according to a six-way broad phonetic classification and then using pattern-matching to compare the stimulus with all the lexical entries matching this specification. In this way they were able to measure the size of 'equivalence classes', that is, the number of words in the lexicon agreeing with the featural specification of the input. Huttenlocher and Zue used the same approach to examine the effect of restricting the analysis to stressed syllables, with markers to indicate the location of unstressed syllables. They found that the performance resulting from this partial specification approach was not dissimilar to that obtained when the word was specified in full: for example, the mean equivalence class in the full specification case was 210, compared with 291 in the partial specification case. In a third condition, they examined the effect of specifying unstressed syllables, leaving markers for stressed syllables, and found this approach far less efficient, with an average of 3717 words matching the input in the lexicon.

Huttenlocher and Zue's finding seemed particularly encouraging because of the apparent robustness of manner features, both in terms of the tendency of speakers to preserve the manner of articulation of segments in otherwise degraded utterances and of the perceptual mechanisms utilised by hearers, who rarely misperceive the manner of articulation of consonants (Miller and Nicely 1955). However, Carter (1987) has cast doubt on the relative superiority of stressed syllables in partitioning the lexicon. He used four different approaches to transcribing the words in a frequency-weighted lexicon: phonemes, manner features, place features, and a 'null' condition in which only the number of phonemes was specified. Information about stress pattern was also encoded. Although the condition in which stressed syllables were transcribed phonemically and unstressed syllables were not transcribed was more effective than the alternative condition in which phonemic transcriptions were provided for unstressed syllables only, the other types of transcription revealed no such superiority for stressed syllables. Carter therefore questions Huttenlocher's (1985) claim that automatic recognition algorithms could be made more robust by exploiting the manner information encoded in stressed syllables. It is not clear, however, whether Carter's criticism extends to the human word recognition mechanism; although Miller and Nicely's 1955 study revealed hearers' dependence on manner cues when they heard speech in noise, in the normal case they may have access to a description of the input which approximates more closely (though clearly not exactly) to a phonemic transcription for stressed syllables (though not necessarily for other parts of the word), and hence may be able to utilise the superior information provided by stressed syllables in this case.

Thus it may indeed be the case that for human hearers, stressed syllables are more informative than unstressed syllables. Altmann and Carter (1988) have shown that the enhanced informativeness of stressed syllables in analyses like that conducted by Huttenlocher and Zue arises from the fact that in such syllables a wide range of phonologically distinct vowels may occur (i.e. any full vowel), while the range of vowels occurring in unstressed syllables is relatively small. Stressed syllables are thus a more efficient means of partitioning the lexicon.

2.2.2. Word onsets and phonological reduction

As Section 2.2.1. shows, there is a considerable body of evidence suggesting that the acoustic-phonetic material present in stressed syllables ought to offer relatively reliable cues to the identity of words. For this reason, models of word recognition in which stressed syllables play a central role have been proposed (see Section 4). However, for reasons which will be explored in Sections 3 and 4, many writers have proposed models in which lexical access is initiated via analysis of the initial portions of words. This seems an attractive notion in the sense that lexical access via initial syllables would tend to lead to faster recognition than access via other parts of words, since by their definition initial syllables occur earlier in the word than other syllables. Unfortunately, however, there is only scant evidence that word onsets provide similar islands of reliability in the speech signal.

The studies of phonological reduction discussed above (Brown 1977, Gimson 1982 etc) make no mention of the reliability of word onsets: phonological modifications apply just as frequently to initial syllables as to other parts of the word. However, Cutler et al.(1985) have identified a tendency on the part of languages to preserve the structure of word stem onsets in a number of ways. For example, there is a tendency for languages to adopt suffixation rather than prefixation, even when other, independent structural characteristics would suggest that the language would be more likely to adopt prefixation. Similarly, they argue that the results of Cooper and Paccia-Cooper (1980), which were discussed above, indicate an awareness on the part of the speaker that word onsets are a functionally important part of the word as far as the hearer is concerned; they note that it is only the frequency of the word *following* the boundary (i.e. the word whose *onset* would be affected) which is implicated in the blocking of palatalisation. While this is a very plausible account in some respects, it does not explain why word onsets are not fully immune to most, if not all, phonological processes, as are stressed syllables.

It should be noted, however, that while stressed syllables and word-initial syllables are logically distinct entities, in reality the two frequently coincide in English. Cutler and Carter (1987) examined the metrical characteristics of words in two computerised dictionaries and in a corpus of spontaneous speech. They discovered that a high proportion of words in English bear stress (primary or secondary) on the initial syllable: for example, 69% of content words in the MRC Psycholinguistic database (Coltheart 1981) and 75% of the content word tokens in the London-Lund corpus (Svartvik and Quirk 1980) fell into this category. However, the question of whether it is initial syllables *per se* which are used to access the mental lexicon is still a valid one.

In summary, then, although there is a limited case to be made, based on the linguistic evidence, for the immunity of word onsets to phonological and other sorts of modification which might be expected to interfere with the lexical access process if word onsets were important in this process, the case in favour of the central role of stressed syllables is far stronger.

2.2.3. Summary of evidence relating to variability

The evidence cited in Section 2.2. indicates that hearers cannot rely upon invariance in the acoustic signal: the surface realisation of words will vary from utterance to utterance, depending on a wide variety of factors such as the identity and emotional state of the speaker, and the style and rate of the speech produced. However, the variability which besets the acoustic-phonetic level is not as unconstrained as it might at first appear. First, certain parts of words (stressed syllables, and to a lesser extent, word onsets) are relatively immune to phonological modification. Second, higher-level factors such as syntactic structure, word frequency, sentence topic and redundancy appear to operate in a systematic manner to govern the application of phonological rules.

2.3. Summary of linguistic evidence

This section set out to characterise the input to listeners, with a view to determining those aspects of fluent speech which might be expected to cause difficulties for the recognition process. A first problem which was identified arose from the lack of reliable word boundary cues in fluent speech. In casual conversation, word boundaries are unlikely to be marked on a systematic basis by identifiable acoustic events.

A second problem inherent in fluent speech is the variability of word tokens. Inter-speaker differences arising from physiological characteristics, and intra-speaker differences due to the mood and state of health of the speaker, as well as to long-term settings of the vocal apparatus, interact with other sources of variability such as speech rate, style and phonological context, to produce a potentially infinite variety of word tokens representing a single word type. Although the phonological processes are constrained to some extent by higher-order factors such as syntactic structure and word frequency, it is evident that the variability of pronunciations of words is a phenomenon which has non-trivial implications for hearers and for models of word recognition, in terms of the difficulties it will cause for selecting a set of word candidates. However, variability is not evenly distributed across words: stressed syllables are major 'islands of reliability', preserving a relatively invariant relationship between tokens. The value of stressed syllables in this respect is likely to increase in conditions of overall unintelligibility, as might be encountered, for example, in conversational speech.

3. Perceptual studies

The studies discussed in the previous section pointed to two potential sources of difficulty for the hearer engaged in the process of word recognition: first, no consistent word boundary cues could be found, and second, the relationship between word tokens representing single types was indirect. Together, these suggest that hearers may have difficulty in achieving a phonological mapping between input and lexicon. However, researchers' failure to identify word boundary cues and to perceive the unity of related word tokens need not necessarily reflect a similar inability on the part of hearers. The human word recognition mechanism might be designed in such a manner that these aspects of the speech signal cause no impediment to word recognition. To assess the effect of these characteristics of the speech signal on the hearer, we must turn to perceptual studies of the word recognition process.

Evidence arising from perceptual studies of auditory word recognition will be presented in this section. As we shall see, the lack of segmentation cues and the variability in the speech signal do indeed present impediments to efficient phonological mapping between input and lexicon. The models of word recognition presented in Section 4 will need to account for these findings, as well as the use by hearers of contextual information to improve the efficiency of word recognition, which has also been demonstrated in studies discussed below.

A number of techniques have been devised to investigate the mechanisms involved in auditory word recognition (6). Some tasks simply require the subject to report the word which has been heard. Stimuli may be presented in good hearing conditions or in noise. A further technique which should perhaps be grouped with these naming tasks is gating, in which subjects hear a spoken stimulus presented in increments (words or subword fragments) and must report what they believe the whole stimulus to be. Other tasks involve listening to an auditory stimulus and pressing a button when a pre-specified target is perceived. In these monitoring tasks the targets include phonemes, whole words (which may be specified by indicating the word itself or one of its attributes) and mispronunciations. A further paradigm, which has been used extensively to examine the processing of ambiguous items, is cross-modal priming: subjects are asked to press a button to indicate the word/nonword status of a visual stimulus while listening to a sentence which may or may not contain a related word. Finally, the shadowing task also involves two modalities: this time, subjects are asked to report aloud, and at very short delays, auditorily presented messages which may contain errors.

3.1. Segmentation

In Section 2.1, a number of word boundary cues were identified. Nakatani and Dukes (1977) listed several such cues and also reported that in an identification task their subjects were between 33.3% and 100% successful at distinguishing sequences such as *no notion* and *known ocean* on the basis of such information. Similarly, Lehiste's (1960) subjects correctly recognised potentially ambiguous phoneme sequences between 25% and 100% of the time.

There is thus some evidence that hearers can make use of segmentation cues -- when they are present. However, it was noted in Section 2 that according to Brown (1977), Gimson (1982) and Vaissiere (1985) word boundary cues were often absent in all but the most formal styles of speech. The fact that the lack of segmentation cues in the speech signal causes problems for hearers is evidenced by a number of studies, both observational and experimental. Bond and Garnes (1980) examined a corpus of approximately 1,000 misperceptions of fluent speech in order to determine the kinds of processes which were involved. They found that, although many of the errors involved such phenomena as segmental substitutions and phonological errors, 17.9% resulted from hearers' failures to identify correctly the location of the word boundary. These comprised boundary deletions (8.3%), boundary shifts (4.4%) and boundary insertions (5.2%).

Several studies have shown that the behaviour of hearers with respect to segmentation adheres to certain constraints. Taft (1984) presented listeners with sequences which could be parsed as either one or two words (e.g. /ˈlɛtəs/ --> *lettuce* or *let us*; /ɪnˈvɛsts/ --> *invests* or *in vests*; hearers were more likely to select a two-word reading when they heard a sequence with an unstressed initial syllable (e.g. /ɪnˈvɛsts/) than when they heard a sequence with initial stress.

Cutler and Norris (1988) noted that Taft, like others, had confounded the stress and vowel quality variables. In their main experiment, they sought to determine the effect of differing vowel quality on segmentation. They distinguished strong syllables, which contain full vowels, from weak syllables, with reduced vowels. They presented hearers with disyllabic non-words and asked them to monitor for any which started with a real word. Examples of such items were *mintayf* (syllabic structure: strong-strong) and *mintesh* (syllabic structure: strong-weak). Note that in these examples, which are representative of the materials used, the syllable boundary falls within the real word, i.e. after the /n/ of *mint*. They found that reaction times to real words such as *mint* were longer when the nonword had the strong-strong structure. They interpreted this result as evidence of a segmentation strategy which used strong syllables to hypothesise word boundaries; when two strong syllables are adjacent and the acoustic material representing one of the words overlaps with the other, an interference effect results in longer reaction times.

Butterfield and Cutler (1988) reported an observational and an experimental study. They first examined the errors published in 'slip of the ear' studies such as that of Bond and Garnes as well as a corpus collected by Cutler. Of these, they selected the 139 instances of word boundary misperceptions which involved the misplacement of the word boundary across a syllabic nucleus (e.g. *won't bother me --> lobotomy*). They found that there was a tendency for word boundaries to be inserted before syllables containing a full vowel and deleted before syllables containing a reduced vowel; such cases accounted for 95 of the 139 tokens in their corpus. This result was not due to a preference on the part of the hearers to report higher-frequency words than those actually uttered; nor does it seem likely that the result was due to greater opportunity for hearers to insert erroneous boundaries before syllables with full vowels, since Cutler and Carter (1988) had shown the reverse to be the case -- about 75% of syllables containing full vowels in a spontaneous sample of English speech were already word-initial.

Noting that the studies of naturally-occurring slips-of-the-ear reported above are beset by various methodological difficulties, Butterfield and Cutler sought to induce perceptual errors in subjects by presenting word sequences at low amplitude levels and noting the word sequences which hearers reported hearing. Their results were similar to those obtained in the observational study: of the misperceptions, the 168 responses which preserved the number of syllables of the original contained 257 word boundary misplacements, and of these, 196 (76%) were either boundary insertions before syllables containing full vowels or boundary deletions before syllables containing reduced vowels.

Cutler and Norris (1988) have proposed a strong syllable segmentation strategy to account for their own findings as well as those of Taft and of Butterfield and Cutler. The strategy they propose involves the insertion of a tentative word boundary at the onset of any strong syllable and the initiation of lexical access attempts at such points. In Cutler and Norris' terms, 'strong syllable' is defined as above, that is, as a syllable containing a full vowel. This interpretation includes not only stressed vowels, but also unstressed and unreduced vowels (as in the first syllables of *monsoon* and *typhoon*). However, in order to identify the phonological category of full vowels in the speech stream, the hearer needs to perform not only segmentation but also classification of the input. Norris and Cutler (1985) have argued that while classification of this kind requires segmentation, the reverse is not true: it is possible to segment the speech stream without attempting to provide a classification in terms of linguistic units. Should one wish to construct a model in which segmentation preceded classification, it would be necessary to identify some acoustic feature which would correlate well with vowel quality, so that when the hearer encounters this feature in the speech stream and makes a segmentation, the resulting segment will tend to correspond to some useful higher-level unit of processing. Cutler and Norris (1988) propose that an appropriate feature in this instance is a combination of vowel duration and other features.

> ... segmentation at strong syllables could be incorporated into a model involving no classification at all ... Suppose that the segmentation device simply monitored the incoming waveform for high-energy quasi-steady-state portions of a certain minimum duration ... ' (7)

Cutler and Norris point out that such a strategy would frequently be successful; Cutler and Carter (1988) have demonstrated that syllables with full vowels are often found at the beginnings of words in English (and vowel length and quality correlate because of the related phenomenon of stress). In order to explain their own experimental results, they reason that, if hearers were following the strong syllable segmentation strategy, the real word would be detected more slowly in the strong-strong case than in the strong-weak case, since in the former, another segmentation (and hence another lexical search) would be instituted at the onset of the second strong syllable. Delays would result either because of the conflicting processing demands of the two sets of searches or because of the need to reassemble phonetic material distributed across two syllables. Results of the main experiment were consistent with this claim. However, in a control experiment, no such difference between the strong-strong and strong-weak condition should be found in pairs such as *thintayf* and *thintef*, because the real word would not straddle a syllable boundary in these examples. Their results supported this account (8).

On the basis of the perceptual studies presented so far, it seems that hearers cannot often determine, pre-lexically, the location of word boundaries with sufficient accuracy for the beginnings of words to be identified in every case. However, despite this evidence, Cole and Jakimik (1980:149) claim that hearers are indeed able to locate word boundaries, by the simple expedient of efficient, word-by-word processing of utterances:

> we suggest that, in perception of conversational speech under good listening conditions, recognition of one word is complete before recognition of the following word begins.

In other words, Cole and Jakimik suggest that hearers know where word 2 begins because they know where word 1 ends. Unfortunately, Cole and Jakimik's premise does not withstand empirical testing, as Bard et al.'s 1988 study, discussed in the next section, shows: in conversational speech, approximately one word in five fails to be recognised by its acoustic offset, a level of failure which appears too high for the system Cole and Jakimik outline to tolerate. Note that the task of Bard et al.'s subjects should in some ways have been easier than that of word recognition under normal circumstances, since in their experiment word boundary information was explicitly supplied in the form of gating increments.

The studies reported in this section have addressed the segmentation issue. The evidence which they present is clear: first, as expected, the lack of systematic cues signalling the location of word onsets does cause problems for hearers. Second, in the absence of such cues, hearers nevertheless behave in a manner consistent with the adoption of a systematic segmentation strategy in speech processing.

3.2. Variability

The second potential problem which was identified in Section 2 concerned the variable nature of the speech signal: the tokens representing a single word type were subject to the modifying influence of a wide range of factors which included the age, sex, mood and state of health of the speaker, the rate and style of the utterance, and the effects of phonological context in which the word was uttered. While it was found that the phonological processes were constrained by higher-level factors such as syntactic structure and word frequency, it might be expected that, even if hearers were aware of the relationship between acoustic and higher-level factors (as they apparently are: see Scott and Cutler 1985), such variability would tend to increase the difficulty of the word recognition task for the hearer. The evidence of perceptual studies discussed below is that this is indeed the case.

One of the earliest investigations of the variability of word tokens was conducted by Lieberman (1963), who had speakers read pairs of sentences in which an identical test word was presented in either a highly constraining or a less predictive context (e.g. the word *nine* was preceded by either the highly predictive sentence fragment *a stitch in time saves* or the less constraining *the number you will hear is*). When the test word was excised from context, hearers found the token excised from the highly predictive context less intelligible than the token excised from the unconstraining context. This finding, which was replicated by others (e.g. Wheeldon 1985; Hunnicutt 1985) was explained by Lieberman in terms of a tradeoff between the information content of the sentential context and that of the word token; speakers, he suggested, adjusted the information in individual words in the sentence in relation to the syntactic and semantic information in the sentence in general. Pedlow and Wales (1988) dispute this explanation. When they conducted a similar experiment, with words heard in normal sentence, anomalous sentence and list contexts, they found no intelligibility difference between conditions. They sought to explain this discrepancy between their results and those of other researchers in terms of the manner in which the materials were produced by the speakers: Lieberman (1963), Wheeldon (1985) and Hunnicutt (1985) blocked the materials according to context type, while the speakers in Pedlow and Wales' first experiment produced the normal, anomalous and list stimuli in an interleaved manner. This explanation is supported by their subsequent experiments. Pedlow and Wales conclude that in adjusting the information value of word tokens speakers are sensitive to linguistic structures larger than the sentence.

Fowler and Housum (1987) studied the effects of word attenuation on hearers' recognition performance. Having identified word tokens which represented the first and second mentions of words in a monologue, and having discovered that speakers tended to shorten these second (or 'old') tokens relative to the first ('new') tokens, they presented the excised word tokens to hearers. They found that hearers made more errors on the old (i.e. shorter) tokens than on the new (longer) tokens, and that the rate of recognition of old words correlated positively with the degree of attenuation of the old token relative to the new. Thus the variability between word tokens on an acoustic level is mirrored in the performance of hearers: the more the token departs from the idealised representation of the word (to which the new token presumably approximates), the more difficult is the recognition task.

Like most other researchers in the word recognition field, Fowler and Housum used read rather than spontaneous materials in their experiments. Exceptions to this general trend are Pollack and Pickett (1963), Wheeldon (1985), Bard et al. (1988) and Mehta and Cutler (1988), all of whom examined the processing of spontaneous speech. As the studies discussed in Section 2 demonstrated, spontaneous or casual speech is more vulnerable than carefully articulated speech to phonological modification. The consequences of this are apparent in perceptual studies of casual speech.

Wheeldon (1985) compared the intelligibility of the same word types produced by the same speaker in three different speech styles: in lists, in spontaneous conversation, and in read speech. After recording the speaker in a conversational setting, she selected test words randomly, then had the same speaker record the same words in their original contexts, but this time reading from a transcript of the selected conversational utterances. Finally, the speaker produced the test items embedded in randomly-ordered lists of words. Wheeldon's results indicate that the variability which is apparent when different tokens of a single utterance are analysed by the linguist also has an impact at the perceptual level. In a first experiment, the three types of stimulus were presented to hearers who were to write down what they heard. While the list and read utterances reached a moderate level of intelligibility (65% and 45%, respectively), the rate of identification of spontaneous utterances was very low, at 6.5%. In a second experiment, utterances were heard in the contexts in which they were originally produced, using the gating paradigm in which stimuli are presented in increments (in this case, word by word) and subjects report what they believe the utterance to be at each successive pass. This time, the read and list tokens were highly intelligible even on first presentation (94% and 82%, respectively) while the spontaneous utterances attained only 40.6% intelligibility on initial presentation. Clearly, then, the acoustic-phonetic differences between word tokens intended to represent the same word type have repercussions at the perceptual level, whether the word is heard in context or in isolation. In particular, words produced in casual, spontaneous speech are difficult to recognise even when they are heard in context.

Wheeldon's results indicate that although the failure to recognise words is likely to increase in spontaneous speech, words may not be recognised until after their acoustic offset even in read speech. Grosjean (1985) examined the phenomenon of 'late recognition' of read speech stimuli in more detail. He used a variant of the gating paradigm in which stimuli are incremented in sub-word fragments to investigate the recognition of low-frequency words after their acoustic offsets. He found that words often failed to be identified until some portion of the subsequent word had been heard, and that subjects' confidence in their response was similarly low until the presentation of further lexical items.

It is in spontaneous speech, however, that the late recognition phenomenon is most frequently observed. Pollack and Pickett (1963) presented to hearers word-gated fragments of conversational utterances which they had surreptitiously recorded. They found that words in early positions in sentences often failed to be recognised on their initial presentation, and only became intelligible with the addition of subsequent context. More recently, Bard et al. (1988) have reported a large-scale study of late recognitions in spontaneous speech, and have concluded that the phenomenon is so widespread -- accounting for 21% of all successful recognitions -- as to warrant serious consideration in models of word recognition. They dispute the claims of writers such as

Marslen-Wilson et al.(1978) and Cole and Jakimik (1980) that word recognition is a left-to-right process, that hearers can know where words begin because they know where the preceding word ends, and that the recognition of one word is complete before the recognition of the following word begins. Tokens which were recognised late tended to share certain characteristics: they were short, often near the beginning of utterances, and frequently belonged to the functor form classes. Finally, the likelihood of a token's being recognised, if hearers failed to do so on initial presentation, increased with the amount of subsequent context available. Subsequently, Shillcock et al.(1988) were able to define further the conditions which were likely to contribute to a word's recognition: metrical stress and phonological phrase boundaries increase independently the probability that a word will be recognised by its acoustic offset, although only phonological boundary strength is associated with the number of post-offset recognitions of previously unrecognised words. Interestingly, when the presence of strong initial syllables was taken into account in the analysis, it was found to explain the apparent effect of the functor/contentive distinction in predicting immediate recognition, although the latter variable was still implicated in the prediction of late recognition points.

Mehta and Cutler (1988) used the phoneme monitoring task to investigate the processing differences between read and spontaneous speech. Unlike Wheeldon (1985), who had used naming and gating tasks, they found no overall difference between the two speech modes: reaction times to word-initial phoneme targets in read sentences were not reliably different from those collected when subjects heard identical sentences produced spontaneously. However, when the materials were analysed with respect to a number of variables known to affect phoneme monitoring reaction times, a more complex pattern emerged. Factors affecting phoneme-monitoring reaction time which had been identified in earlier studies using read speech included the target-bearing word's transitional probability, the length of the preceding word, the position of the target in the sentence, and the accentedness and lexical stress pattern of the target-bearing word. In Mehta and Cutler's study, only transitional probability exerted the same effect in both speech modes (with targets on highly predictable words eliciting faster reaction times than those on unpredictable words). For two of the remaining variables, the effects found in earlier studies were replicated for read but not for spontaneous speech: preceding word length and position in sentence did not affect reaction times in spontaneous speech, but in read speech longer preceding words and later sentence positions were associated with faster responses. By contrast, for the two remaining variables, sentence accent and vowel quality, the pattern of results found in earlier studies was replicated for spontaneous but not for read speech: accented target-bearing words and target-bearing syllables with unreduced vowels resulted in faster reaction times.

Mehta and Cutler explain their results in terms of the differing prosodic structure of read and spontaneous speech. They suggest that the absence of the preceding word length effect can be explained by the presence of pauses immediately preceding many of the test words which were preceded by monosyllables: although targets preceded by monosyllables ought to result in relatively long reaction times, the presence of a pause would enable the subject to complete processing of the preceding word more efficiently than in fluent speech. To explain the lack of the sentence position effect, the authors point to the explanation proposed by Shields et al.(1974), who suggested that, in their own read materials, later targets allow speakers to utilise the predictive properties of intonation contour. But in spontaneous speech, argue Mehta and Cutler, prosodic units are shorter and are less likely to correspond to the sentence as a whole, and therefore the sentence position variable is less likely to correlate with reaction time. For the two variables which exhibited effects for spontaneous speech but not for read tokens, they also advance explanations in terms of prosodic structure. They attribute the presence of a sentence accent effect in Cutler's (1976) experiment to the fact that her test sentences led the speaker to produce contrastive stress, a phenomenon which did not appear in Mehta and Cutler's read sentences. They suggest, therefore, that in the read condition of this experiment, there were fewer opportunities for their subjects, compared with those of Cutler (1976), to use prosody predictively; but the more varied prosodic structure of spontaneous speech made accent a useful feature. This interpretation is consistent with the findings of Tyler and Warren (1987) who used the word monitoring task to investigate the importance of different kinds of structure in the processing of spoken words; they found that interfering with the prosodic structure, particularly within phonological phrases, had an adverse effect on processing. Finally, the syllable stress effect for spontaneous but not for read speech was explained in terms of the value of stressed syllables in terms of 'islands of reliability' in spontaneous speech. Mehta and Cutler's overall conclusion, therefore, is that the manner of processing does not differ between read and spontaneous speech; what does differ, however, is the extent to which hearers have the opportunity to use prosodic cues.

The studies discussed in this and the previous section show that the lack of segmentation cues and the variability of the realisations of words do cause problems for hearers to the extent that they are often unable to recognise a word before its acoustic offset; yet they are still, apparently, able to accomplish a mapping between the auditory input and the phonological representation in the lexicon. Two candidates -- stressed syllables and word onsets -- have been proposed for the unit of processing which is used to accomplish this mapping, and evidence relating to these units will now be presented.

3.2.1. Stressed syllables

Evidence presented in Section 2.2.1. suggested that stressed syllables should provide a reliable source of information for the hearer, a fact that had led to speculation that hearers must make particular use of stressed syllables in recognising words. To see whether this is indeed the case, we turn now to perceptual studies of the role of stressed syllables in word recognition.

There is copious evidence that stressed syllables are in some sense easier to process than unstressed syllables. For example, Bond (1971) found that hearers could more accurately locate clicks in stressed than in unstressed syllables, while Browman (1978) found that stressed syllables were far less susceptible than unstressed syllables to being misinterpreted by hearers. Cole et al. (1978) found that listeners could more easily detect mispronunciations in stressed than in unstressed syllables.

Various researchers have found enhanced phoneme-monitoring reaction times to stressed syllables (Shields et al, 1974; Cutler and Foss, 1977). Shields, McHugh and Martin (1972) embedded nonsense words (masquerading as proper names) in sentence contexts and had subjects monitor for their word-initial phonemes. They found a decrease in reaction times when the target-bearing initial syllable was stressed, compared with the condition in which it was unstressed. However, in a control condition in which the nonsense words were excised from their sentence contexts and presented in strings of other nonsense words, phoneme monitoring reaction times were unaffected by the stress level of the target-bearing syllable. The authors concluded that, although the presence of sentence stress clearly exerted some effect on word-level processing, that effect could not reside in the acoustic characteristics of the signal itself, since this would predict that the stressed syllable advantage would persist even in the control condition. Instead, they ascribed the effect to the rhythmic structure of the sentence, which would lead hearers to expect a stressed syllable at a particular location; this predictive mechanism would of course be absent in the control condition.

Cutler (1976) pointed out that that the task in Shields et al.'s control experiment (monitoring for targets in nonsense strings) may have been so alien to normal speech processing activity that the results might not be reliable, and that perhaps, after all, the acoustic and phonological properties associated with stressed syllables could make some contribution to intelligibility. She therefore constructed an experiment which tested the explanation proposed by Shields et al, this time using real word targets. She produced pairs of test sentences whose members contained an identical word token, but in one version the preceding intonation contour cued sentence stress on the word, and in the other it led the hearer to expect stress on a later word. These spliced sentences, as well as the unspliced originals, were presented to hearers who were to monitor for phoneme targets which occurred initially on the test word. For both the spliced and the unspliced sentences, reaction times were shorter when the preceding intonation contour predicted stress on the item than when it did not (regardless of whether or not the word was *actually* stressed), indicating that hearers were indeed using rhythmic information to predict the positions of upcoming stress. The difference between the spliced and unspliced conditions did not reach significance, but Cutler notes that the magnitude of the improvement in the predicted-stressed compared with the predicted-unstressed condition was three times greater in the unspliced condition than in the spliced condition, a finding which she attributes to the differing contributions of acoustic information. Cutler's conclusion, on the basis of the 1976 paper, was that although the acoustic characteristics of sentence stress have some role in word recognition, more importance should be attached to the predictive value of the rhythmic structure.

In other papers, Cutler and her colleagues pursue the issue of the reason for hearers' apparent use of this attention-directing mechanism. That is, why should they wish to direct their attention to the locations of sentence accents? Cutler and Foss (1977) investigated the possibility that the value of predicting accent might be that it usually, though not always, fell on content words, which in turn might be expected to be more useful in terms of information content. They examined this hypothesis by constructing phoneme monitoring stimuli in which the target word was either a content word or a function word and either bore accent or did not. They found that the presence of accent facilitated reaction time across both grammatical groups, but the contentive/functor distinction was irrelevant. They therefore concluded that this form class explanation could not account for the tendency of hearers to use the predictive force of preceding intonation contour. Cutler and Fodor (1979) indirectly examined an alternative explanation: perhaps hearers used intonation contour to identify the location of sentence accents because this would usually correspond to the location of focussed words in the sentence. Since English has a number of mechanisms for signalling focus, if the explanation were correct, effects consistent with this general account should be found in all locations of focus. Cutler and Fodor chose to examine the focussing phenomenon by presenting a single sentence token containing a phoneme monitoring target, preceded by one of two questions. In one case, the lead-in question focussed the target word, and in the other, it focussed a different word. Cutler and Fodor found that phonemes which appeared in focussed words were responded to faster than phonemes in unfocussed words. It thus seems likely that the reason for hearers' use of preceding intonation contour as an attention-directing mechanism is that, like other mechanisms, it too will tend to lead them to focussed words (9).

Other studies have revealed a detrimental effect on word recognition of distorting stress and/or its associated vowel quality. Bond (1981) played subjects recordings of paragraphs in which voicing, nasality, place and vowel quality contrasts (stressed vowels only) had been manipulated, and compared their performance on two measures, recall and shadowing, with that obtained when they heard undistorted paragraphs. She reasoned that (1981:90)

> altering those properties which are perceptually most salient would be expected to lead to maximal disruption of perception and/or comprehension of the utterance.

She found that manipulation of stressed vowels caused a decrement in performance, across both measures, and inferred from this result that 'the contribution of stressed vowels to the perception of continuous speech is great'.

Others have found similar effects associated with the disruption of stress pattern. Bansal (1966) examined the effect on English hearers of the unusual stress patterns which Indian speakers of English adopt, and found that such forms were extremely detrimental to intelligibility. Interestingly, hearers often reported hearing words which corresponded to the stress pattern produced by the speaker rather than its segmental characteristics, which seems to suggest that the prosodic characteristics of words can override segmental features as far as listeners are concerned. Lagerquist (1980) found that puns which altered the stress patterns of words were unsuccessful.

Cutler and Clifton (1984) examined the role of lexical stress information in the recognition of isolated words or of words in short utterances. The first in the series of experiments which they reported asked whether providing prior information about the stress pattern of words could enhance recognition, as we might expect if stressed syllables were so important to the word recognition process: they presented test words of the structure stressed-unstressed (e.g.*tiger*) or unstressed-stressed (e.g. *canoe*), and measured reaction times. They found that although words with unstressed initial syllables were recognised more slowly than words with stressed initial syllables, words presented in 'pure' lists (containing words with the same stress patterns) were recognised no faster than words presented in 'mixed' lists (differing stress patterns). Thus, it appears that prior knowledge of stress pattern does not aid identification.

In a second experiment, Cutler and Clifton tested the hypothesis that hearers are able to utilise the relationship between the stress pattern and grammatical category of words in English: disyllables with initial stress are usually nouns, while disyllables with second-syllable stress are usually verbs, although counter-examples can be found. They presented subjects with all four types of stimulus, preceded by either *to* (which, followed by the uninflected form, would indicate a verb) or *the* (which legally precedes nouns). The subjects' task was to judge whether the sequence was grammatical or ungrammatical (e.g. *to conTRAST* and *the CONtrast* are grammatical, whereas *to*

CONtrast and *the conTRAST* are not). The results indicated no facilitation of reaction times to words with appropriate stress pattern, and no significant interaction between acceptability, part of speech and stress pattern. Moving stress in a 'leftward' direction (i.e. from the second to the first syllable) was less detrimental than rightward stress shift; Cutler and Clifton explain this in terms of the occurrence of this phenomenon in normal English speech, in which leftward stress shift (as in the well-known *thirteen men* example) is not uncommon. Once again, they found that words with stressed initial syllables were responded to more quickly than words with unstressed initial syllables.

In their final experiment, Cutler and Clifton asked whether correct stress pattern is important for recognition, by manipulating the stress pattern of stressed-unstressed and unstressed-stressed disyllables. They noted that stress changes were often accompanied by vowel quality changes (as in forms such as *subject* which may be a noun /ˈsʌbdʒɛkt/ or a verb /səbˈdʒɛkt/), and that most studies which claim to examine stress in fact confound the stress and vowel quality variables. Two types of stimulus were selected: either both syllables contained full vowels (e.g. *nutmeg, typhoon*), or the unstressed syllable contained a reduced vowel (e.g. *wisdom, deceit*). For each test word, two stimuli were produced: one with the correct stress pattern, and one in which the stress pattern was reversed (in words with schwa in the unstressed syllable, the corresponding vowel produced in the stress pattern reversal condition was an unreduced counterpart). Subjects were asked to judge whether or not the word they were hearing had a physical referent, and reaction times to this decision were collected. The results supported the hypothesis that correct stress pattern was important for word recognition: although the vowel quality variable exerted a strong influence on recognition, reaction times were faster overall for items with correct stress pattern than for those with reversed stress pattern. Bond and Small (1983) found a similar result. Thus, the results of Cutler and Clifton's experiments do not provide an unequivocal view of the role of lexical stress in word recognition: while alterations to the stress pattern of words hinders recognition, other effects which might have been expected were not found.

It was noted in the introduction to this chapter that the word recognition process was customarily described as consisting of two sub-tasks, namely, lexical access and lexical retrieval. Although the effects just described are clearly associated with the word recognition process at some level, and appear to indicate that stressed syllables are easier to process than unstressed syllables, the studies cited do not allow us to determine the precise location of the effects. They may be a function of the access process or of the retrieval process; indeed, they may arise from neither, for, if Cutler and Norris (1988) are correct, hearers operate a prosodically guided segmentation strategy which may itself give rise to the appearance of a word recognition process driven by stressed syllables.

Cutler (1986) has attempted to determine whether there is evidence for the use by hearers of lexical stress information at the lexical access stage. She carried out a set of experiments in which the effect of lexical stress on lexical access was isolated from the effects of vowel quality. Cutler noted that if lexical stress information is *not* used at the lexical access stage, the lexical representations of the two members of pairs such as *FORbear* and *forBEAR*, which are segmentally identical but differ in stress pattern, will be the same. The cross-modal priming task has been used to examine hearers' processing of homophonous items. This task involves hearers deciding whether a visually presented stimulus is a word or a nonword while they are listening to a sentence which may or may not contain a word related to the visual target. Swinney (1979) used cross-modal priming to establish that a visual stimulus presented immediately after the offset of an auditorily presented homophone is facilitated if it is related to any sense of the homophone (not just the sense indicated by context). The effect is transient, however; if the visual probe is presented less than a second later, it is facilitated only if it is related to the contextually-appropriate meaning of the homophone. Cutler reasoned that, if the lexical representation of the words *FORbear* and *forBEAR* ignored the stress difference, and the items were thus homophonous, the cross-modal priming results obtained with such words as stimuli would be consistent with those obtained by Swinney for his homophonous items. She identified eleven suitable items -- apparently the only appropriate pairs in the language -- but after pre-testing had to exclude three pairs. The remaining eight pairs were included in the cross-modal priming experiments. Cutler argued that, if lexical stress information were used at the lexical access phase, then *FORbear* should prime only its associate *ancestor* and *forBEAR* only its associate *tolerate*, while if lexical stress information were irrelevant to the lexical access procedure, the two words should prime either probe. A third possibility, suggested by the results of Cutler and Clifton (1984: see discussion above) was that *FORbear* might prime either probe, while *forBEAR* would prime only *tolerate*. In fact, the results of the first experiment, in which the visual probe was presented immediately after the homophonous word, indicated that the second possibility was correct, and that, at the lexical access stage, *forbear* is a homophone. As in Swinney's 1979 experiment, when the position of the visually-presented item was delayed relative to the offset of the auditory prime, only the contextually-determined meaning produced facilitation. Cutler therefore concludes that lexical access makes no use of stress information, although prosodic information is used at other stages in word recognition (Cutler and Norris 1988).

Van Heuven (1988) has disputed Cutler's claim that lexical stress plays no role in lexical access. Using materials from Dutch (another stress-timed language, like English) he examined the gating and shadowing responses made by subjects who were presented with word-initial fragments of word-pairs which had been matched for the segmental structure of their initial syllables but which differed in terms of the stress level of these initial syllables -- that is, the initial syllables were either stressed or unstressed. After hearing the portion of the word which matched its partner segmentally (but differed in stress level) subjects were offering responses which were accurate in terms of stress level 76% of the time. Van Heuven argued that this result, which he replicated in a shadowing experiment, is evidence that hearers do use lexical stress information in the lexical access process. Connine et al.(1988) have also argued that it is possible to interpret the results of their study (which concerned a lexical bias in the identification of acoustically ambiguous stimuli) in terms of the use by hearers of lexical stress information at the lexical access stage.

Thus it remains unclear whether the stress effects which have been reported are due to the process of lexical access or to some other aspect of speech recognition. While the available evidence suggests that lexical stress information is used at some stage in the word recognition process, the question of whether lexical stress effects can be found at the lexical access stage is a matter of some dispute.

3.2.2. Word onsets

The studies discussed in Section 3.2.2. suggest that hearers find the stressed syllables of words useful cues in the recognition process. This has led to suggestions that stressed syllables are used to initiate lexical access (e.g. Grosjean and Gee 1987 -- discussed below). However, an alternative candidate for the part of the word on which phonological mapping is based is the initial portion of the word (Marslen-Wilson and Welsh 1987; Marslen-Wilson and Tyler 1980; Marslen-Wilson 1987; Cole and Jakimik 1980). This section will consider to what extent such claims are justified by the perceptual evidence.

One method of investigating the perceptual importance of different parts of words which has been widely employed is the listening-for-mispronunciations task, in which subjects are presented with sentences in which certain words have been deliberately altered and are asked to press a button whenever they hear such a word. The dependent variables which are available for analysis in such an experiment are first, the number of occasions on which subjects realise a mispronunciation has occurred, and second, the speed with which they respond to the mispronunciations which they do notice. Cole (1973) was among the first to use this technique. He manipulated both the location of the error and distance, measured in number of articulatory features, between target and mispronunciation: errors were placed on the first, second or third syllable of words, and differed from the target by one, two or four features. The position variable, which is of most interest here, was unfortunately confounded with another factor, in that first syllable mispronunciations involved modification of the syllable-initial consonant, whereas mispronunciations in second and third syllables involved syllable-final segments. Thus, it is impossible to determine whether the result Cole obtained -- faster reaction times to errors involving one distinctive feature in first syllables -- was due to the importance of the position of the syllable in the word or the position of the segment in the syllable.

Other writers, however, have reported results which appear to indicate much more convincingly the importance of the initial portions of words. Nooteboom (1981) examined the issue using a technique other than listening-for-mispronunciations. He selected Dutch polysyllables and identified points at which, both left-to-right and right-to-left, the phoneme sequence was unique. For example, the Dutch word *kiosk* /kijɔsk/ becomes unique at the fourth phoneme (that is, no other word begins /kijɔ/), while /ɔsk/ is the final sequence of no other word. Nooteboom edited single tokens of the test words to yield two stimuli, in which either the first part or the second part had been replaced by a tone: either /kijɔ/ + tone, or tone + /ɔsk/. The subjects' task was to listen to the two stimuli and to provide the full word, or if this was not possible, to repeat the word fragment as accurately as possible. Reaction times were measured. If the initial portions of words are, as some claim, crucial to the lexical access process, removal of them should result in the failure of subjects to recognise the word in question: thus, in the 'tone + /ɔsk/' condition, recognition should be impossible or at least very difficult, while in the '/kijɒ/ + tone' condition, recognition should be readily accomplished because the information after the uniqueness point should be redundant. There was no significant difference in complete failures to perceive the stimulus (i.e. instances where subjects repeated the fragment incorrectly) across the two conditions, suggesting that no part of the word is *so* crucial to the word recognition process that the hearer cannot compensate for its removal. However, more initial portions yielded correct recognitions than final portions, suggesting that word onsets do play an important role in lexical access. On the other hand, reaction times were often considerably longer than those found in other studies when whole words were presented; therefore, the information following the uniqueness point may not have been not completely redundant.

3.3. Context

The studies considered so far in this section have been concerned with the contribution of the acoustic-phonetic properties of words to their recognition. A number of other factors have been shown to affect word recognition performance: in this subsection, the context variable will be considered (10). Much debate has centred around the issue of the use of context in auditory word recognition. There is a considerable body of evidence suggesting that hearers do indeed use context to assist in the word recognition process. As early as 1900, Bagley showed that hearers noticed that a word token had a consonant deleted when it was presented in isolation, but not when it was presented in context. Miller et al.(1951) found that words were more intelligible if they were heard in context. Tasks other than word naming have consistently revealed a context effect. Morton and Long (1976) demonstrated that the presence of predictive context speeded phoneme monitoring reaction times, while Jakimik (1979) found the same result in a monitoring-for-mispronunciations task. Marslen-Wilson (1975) found that subjects' performance on a shadowing task showed evidence of the influence of context on word-recognition; fluent restorations (i.e.

corrections of mistakes planted in the auditory stimuli) occurred more often in a constraining context. Marslen-Wilson and Welsh (1978) also found that listeners' dependence on bottom-up information in a shadowing task varied as a function of contextual constraints. Marslen-Wilson and Tyler (1980) showed that the presence of syntactic and semantic constraints shortened word monitoring reaction times. Grosjean (1980) presented identical gated word tokens in two with-context conditions and in a no-context condition, and found that words were recognised with less acoustic input when context was present than when the token was heard in isolation.

Tyler and Wessels (1983) wished to partial out contextual factors to their syntactic, pragmatic and semantic sources. They constructed, for each of their 25 Dutch test words, four different sentence types defined by the binary variables of semantic constraint (none versus weak) and syntactic constraint (weak versus strong). Since the tokens of the test words could be expected to differ in terms of intelligibility precisely because of their varying predictability (Lieberman 1963) a fifth, neutral sentence was also recorded and the token excised from it was spliced onto the four experimental sentences, as well as being presented in isolation along with the other stimuli in a gating experiment. They found that although syntax exerted a small but significant effect on the amount of sensory input needed for recognition, the influence of semantic constraints was far greater (11). It should be noted, however, that the nature of Tyler and Wessels' materials was such that attending to the syntactic cues would only result in the identification of a verb subclass which might still contain many members, while utilisation of semantic cues would have a much more marked effect in partitioning the lexicon. It is possible, therefore, that their result reflects not processing activity but language structure.

The existence of context effects is widely accepted in psycholinguistics; but their precise locus is still a matter of some contention. That is, while it has been demonstrated that the presence of context contributes to auditory word recognition at some stage, opinions differ as to whether contextual information is used to constrain lexical access or merely to narrow down the set of candidates once accessed. As Section 4 will show, these differing opinions have given rise to distinct models of the word recognition process. Most of the techniques used to investigate word recognition are incapable of resolving the issue because they provide insights only into the recognition process, while the pattern of words accessed prior to recognition is hidden from inspection. One exception to this is the gating paradigm (Grosjean 1980).

Grosjean (1980) used the gating paradigm to examine the context issue. He compared responses offered at early gates when stimuli were heard in long, short and no context conditions, and found that the set of word candidates offered by subjects differed from one context condition to the next. Another dependent variable which Grosjean examined was the point at which the target word began to be offered by at least one subject. The 65 msec mean for the long context condition was significantly lower than the 245msec no-context mean. Furthermore, when the number of candidate types proposed after the first gate (30msec) was examined, the context effect was significant, with the eight subjects who heard the no-context condition offering on average 7.5 different word candidates, compared with 6.08 in the short-context condition and 4.73 in the long-context condition; a similar significant effect was found when the numbers of word candidates offered across all gates in each context condition was analysed. These results were interpreted by Grosjean as evidence of the use of contextual information in lexical access.

Tyler (1984) challenged Grosjean's context result. She examined in detail the data from an earlier gating experiment (Tyler and Wessels 1983), reasoning that, if hearers used context to constrain the selection of the word-initial cohort, then contextually inappropriate candidates should *never* be offered by subjects. Certainly this is true in a strong version of the contextual pre-selection argument; however, such an argument does not take into consideration the possibility that in some cases hearers may, for whatever reason, fail to make use of contextual cues. Such a failure may account for the 5% of contextually inappropriate candidates observed after the first gate in the condition which offered the strongest contextual constraints. Unlike Grosjean (1980), Tyler failed to find any difference in the number of word candidates offered in the isolated and with-context presentations: the twelve subjects who heard any given stimulus offered a mean of 7.1 candidate types after the first gate in the no-context condition, a figure which did not differ significantly from the 7.9 to 8.5 offered in the four with-context conditions in the experiment. Tyler's explanation of the discrepancy between her results and Grosjean's is that the long-context condition in which the significant facilitation was observed 'might just be more effective in reducing the size of a cohort that is activated by the bottom-up input' (1984:421). It should also be noted that the word tokens used in Tyler's experiment would have been maximally distinct because they were excised from a neutral carrier sentence; Grosjean's was a token excised from a highly constraining sentence so it would be expected to be relatively less intelligible (Lieberman 1963). Under circumstances of good intelligibility, such as would therefore pertain in Tyler's experiment, recognition scores even in the no-context condition would presumably be comparatively high, and closer to the with-context condition.

However, a number of recent studies have challenged Tyler's finding. Connine (1988) examined the context issue using materials in which the voice onset time of the stimulus-initial segments had been manipulated. She found that although acoustically unambiguous stimuli (i.e. those taken from the endpoints of the voicing continuum) were unaffected by higher-level factors, acoustically ambiguous stimuli were labelled in a manner consistent with context. Lowe (1988) used the gating paradigm to investigate the issue, and found that hearers can make use of context during access, but that bottom-up information was given priority; when subjects had access to both top-down and bottom-up information, they identified a word-initial segment more accurately than when only one source was present. Lowe's interpretation of the relative contribution of top-down and bottom-up factors is consistent with the results of Connine's (1987) study mentioned above.

To conclude, a wide range of views concerning the role of context in auditory word recognition is to be found in the psycholinguistics literature. Although much experimental evidence exists suggesting that hearers use context in the processing of words, the precise stage at which they do so remains unclear.

3.4. Remarks on the generalisability of results of perceptual studies to other subjects

Before turning from the description of perceptual studies to an exposition of word recognition models, it is useful to consider the extent to which the results reported in these studies can be generalised to listeners other than those who took part in the experiments. Although the statistical significance of the results in the literature reassures us that the findings are indeed representative of those which would be observed if other members of the same population were tested, one might ask whether the subjects who habitually take part in word recognition experiments actually constitute a definable sub-population which differs in important respects from other hearers.

By far the largest group from which subjects are sampled are university undergraduates. While they may be considered a specialised group from a number of perspectives, many of the factors which may distinguish them from other adults (such as age and IQ) have not been shown to exert a consistent influence on language processing abilities. However, one feature which differentiates this group of adults from the rest of the population, and which some authors have suggested may be an important factor in their language behaviour, is their level of literacy. For example, Brown and Yule (1983:14) comment on the effect on spoken language production of protracted exposure to written language, and on

> the distinction between the speech of those whose language is highly influenced by long and constant immersion in written language forms, and the speech of those whose language is relatively uninfluenced by written forms of language.

Linell (1982) has explored in some depth the written language bias in linguistics. He suggests (p. 43) that the written language may exert a strong influence on the spoken:

> It is highly likely that the competence for spoken language of a literate person differs from that of the illiterate ... our phonological intuition may change as a consequence of learning the alphabet.

The limited amount of experimental evidence which is available on this point tends to confirm the suspicions of Linell and of Brown and Yule that literate and illiterate adults use different linguistic processing strategies. Idrissi (1987) compared the performance of Moroccan subjects matched for age and sociological factors but differing in terms of level of literacy on a number of tasks involving the manipulation of phonemes, syllables and words. He found that literacy imposed a view of language structure derived from the knowledge of the spelling of words. Morais et al. (1979) compared the performance of literate and illiterate Portuguese adults on a task which involved adding sounds to or deleting sounds from the beginnings of words and non-words. They found a significant difference in performance on this task depending on the level of literacy attained; however, performance was unaffected by the age at which reading was acquired. This result indicates that awareness of speech as a sequence of phones does not arise spontaneously, but is related to the ability to perform grapheme-to-phoneme correspondences in reading.

Unfortunately, studies relating level of literacy in adults to language performance are very rare. An alternative source of information on this question might be studies in which child performance on language processing tasks is compared with that of adults, if we could be sure that the effects we observed were due to literacy rather than maturation. In this connection, some work in the visual word recognition area is of interest. Davidoff et al. (1982) extended the study of Bruner and O'Dowd (1958) who had presented readers with words whose beginnings, middles and ends were mutilated (e.g. *aviation --> vaiation, avitaion, aviatino*, respectively), and had shown that adults were most hindered in visual word recognition by alterations to the beginnings of words, and least affected by alterations to the middle, with the mutilation of word endings causing an intermediate degree of interference. Davidoff et al. replicated this serial position effect for their highly literate group, but their semi-literate subjects did not show the effect of superiority of word endings over word middles; instead, performance decreased steadily over the three positions (i.e. beginnings > middles > ends). Interestingly, a group of child subjects tested by Davidoff et al. produced results extremely similar to those of the semi-literates. The importance of this point is that it suggests that the potentially confounding age factor may not be an important variable, and that data from the somewhat more numerous child studies can also be taken into consideration in evaluating this issue.

The implications of a similarity in processing behaviour in illiterate adults and children is apparent in a study by Cole and Perfetti (1980), who conducted an experiment using the listening-for-mispronunciations task employed by Cole (1973: see Section 3.2.2.). In their study, however, Cole and Perfetti included child subjects as well as adults. As in other listening-for-mispronunciations studies, the adult subjects produced shorter reaction times to mispronunciations on second syllables than on first syllables. Cole and Perfetti interpret this as evidence for the importance of word beginnings in lexical access. They claim that when a first syllable is altered to produce a mispronunciation, lexical access is hindered, because it is on the basis of this word-initial information that adults retrieve words from the mental lexicon. Instead of responding as soon as the mispronunciation occurs, subjects continue to monitor the acoustic signal until it becomes apparent, when bottom-up analysis rules out all the available candidates, that the error has been made. However, if the mispronunciation is located on the second syllable, access has already been initiated on the basis of the correct prior (i.e. first-syllable) information, and all that is necessary for a mismatch between stimulus and intended word to be detected is a comparison of the two. Thus, they argue, reaction times to first syllable mispronunciations are longer than those to second syllable mispronunciations.

Cole and Perfetti's finding for adults replicates exactly that of the earlier studies (Cole 1973; Cole and Jakimik 1978; 1980a). The children in the study did not, however, display the same pattern of results: in their case, reaction times for first and second syllable mispronunciations did not differ significantly. This discrepancy between adult and child performance led Cole and Perfetti to suggest that children must be following some strategy other than the 'initial syllable' one which they proposed for adults. Cole and Perfetti also note that in reading studies, less skilled readers were more affected by contextual cues than skilled readers, and speculate that the former group uses contextual knowledge to compensate for slowly executing low-level identification processes.

There is indirect evidence, then, that the skills acquired in learning to read can impart specialised abilities in other areas of language processing, which extend even to the auditory domain. Since undergraduates form a sub-population distinguishable from other adults by their enhanced reading skills, we should perhaps be cautious in generalising results obtaining from them to the rest of the adult population. Unfortunately, however, most of the studies which have been discussed in Section 3 have drawn their subjects from just this sub-population.

4. Models of word recognition

The previous two sections have been concerned with the evidence, linguistic and perceptual, which is relevant to the issue of auditory word recognition. The models of the word recognition process which have been developed in response to this evidence will now be considered.

The studies discussed in Sections 2 and 3 of this chapter revealed several rather powerful constraints on the possible form of models of word recognition. First, there is the nature of the input itself: the lack of segmentation cues in the speech signal and the variability in the acoustic form of segments which are considered identical at the lexical level place important requirements on the theories. In particular, they must account for hearers' well-attested abilities to process spoken language, but must not have recourse to claims about the explicit segmentability or invariance of speech. Second, context effects must be modelled: words must be recognised more readily in context than in isolation. A final requirement, which is rarely addressed seriously in the literature, is that the models should be able to function efficiently despite the high degree of unintelligibility of individual words in casual connected speech. In the remainder of this section, we will first consider the claims of the Cohort model (Marslen-Wilson and Tyler 1980; Tyler 1984; Marslen-Wilson 1987). These claims will then be contrasted with those of the Logogen model (Morton 1969) and of the Search model (Forster 1976, 1979), which embody different approaches to the use of contextual cues, and of a group of models which emphasise the importance of parts of the word other than initial syllables in the process of accessing the mental lexicon (Cutler 1976; Gee and Grosjean 1987; Cutler and Norris 1988).

As a starting point in the discussion of word recognition models, we will consider the claims of the Cohort Model (Marslen-Wilson and Welsh 1978; Marslen-Wilson and Tyler 1980; Tyler 1984; Marslen-Wilson 1987). This description will be complicated somewhat by the fact that the theory has undergone revision since its initial appearance. In its earliest instantiation (Marslen-Wilson and Tyler 1980; Tyler 1984) the model was formulated as follows. The recognition of a word was thought of as a two-stage process. The first stage, which Tyler (1984) terms *cohort initiation* and which Marslen-Wilson (1987) terms the *access* phase, involves the assembling of a set of word candidates on the basis of the acoustic-phonetic properties of the input. On presentation of a spoken word, the hearer uses some initial portion of the input (variously described as the word's initial syllable, initial CV, initial 150 msec and initial 200 msec) to access all the entries in the lexicon which match the input for this initial portion. It is important to note that the words which were considered to constitute this candidate set (or *word initial cohort*) matched the phonological, rather than the phonetic specification of the input; thus, for example, the word-initial cohort for the word *stack* includes both *stamina* and *stampede*, even though the vowels in the initial syllables of the latter two words will be affected by the nasal quality of the following consonants, whereas the vowel in the input word would not.

Thus, according to the Cohort Model, the access phase is a strictly autonomous, bottom-up process (Marslen-Wilson 1987:72): only acoustic-phonetic information can be used to assemble the set of word candidates which the hearer considers. The second phase, lexical retrieval (which is also termed *cohort resolution* (Tyler 1984) or *selection* (Marslen-Wilson 1987)), consists of the pruning of the word-initial cohort, using both contextual and continuing acoustic-phonetic information, until a single candidate remains; at this point, the word is recognised. If a word is presented in isolation, its recognition point can be predicted from its uniqueness point, that is, the point at which the word becomes uniquely distinguishable from all other words in the lexicon. For a proportion of words in the language, namely homophonous items, the uniqueness point is not encountered by the end of the word; but it is argued that the majority of words become unique by their acoustic offsets (12). However, a significant number of items fail to do so: Luce (1986) has estimated that some 40% of all items encountered become unique post-offset. When the word is heard in context, syntactic, semantic and pragmatic cues contribute to the process so that recognition may occur well before the uniqueness point. Hearers' ability to make use of the uniqueness point is well supported experimentally. Marslen-Wilson (1980; 1984) reports a series of experiments in which a lexical decision task was used. The non-words used in the experiment varied with respect to the position of their last real-word phoneme (that is, the point at which they became a non-word) relative to the beginning of the word. Lexical decision times were almost constant (around 450 msec) when measured from the uniqueness point.

The 1980 version of the Cohort Model was not without flaws. First, it failed to account for the word frequency effect: recognition time was governed purely by the interaction between the phonological form of the lexical item, its relationship to other entries in the lexicon, and the presence or absence of contextual material. Second, the model predicts that words whose initial portions are mispronounced will never be recognised, because the hearer will assemble the word-initial cohort corresponding to the input rather than the idealised lexical representation of the intended word. In fact, there is abundant evidence, observational and experimental, that this prediction is not borne out. First, as Norris (1982) points out, accidental mispronunciations, such as when a drunk pronounces the word *cigarette* as *shigarette*, do not prevent hearers from recognising words. Second, experiments utilising the monitoring-for-mispronunciations paradigm (e.g. Cole 1973; Cole and Perfetti 1980) demonstrate that although hearers react slowly to words with mutilated onsets, they identify word-initial mispronunciations more frequently than those at other locations in the word, a result which suggests that they can identify words even when the word-initial information is impoverished. Finally, Nooteboom (1981) found that subjects could recognise words even when he removed initial syllables completely and replaced them with non-speech sounds. It might perhaps be objected that these sources of evidence represent the marked case, and that in normal speech production -- unhampered by interference from alcohol or experimenters -- speakers protect word onsets from phonological modification. In fact, as the evidence in Section 2 demonstrates, this is not the case: although word onsets are protected from phonological modification under certain conditions, they are not entirely immune to such effects. Conversely, even when speakers do not alter word onsets in any unusual way, hearers apparently access words which lie outside this phonologically defined cohort (see, for example, Grosjean's 1980 discussion of his gating response data).

In response to the evidence relating both to word frequency and to variability, Marslen-Wilson (1987) modified the theory. The experiments reported in the following chapters were designed in response to the predictions of the 1980 version of the model, but as the results are, where appropriate, also discussed in relation to the later version, it is useful to summarise its characteristics here. The two-stage structure, with an autonomous access phase followed by a selection phase, was preserved:

> early in the word, when only the first 100 - 150 msec have been heard, then the recognition devices corresponding to all the words in the listener's mental lexicon that begin with this initial sequence will become active (Marslen-Wilson 1987:78)

To account for hearer's ability to map imperfect acoustic inputs onto idealised phonological forms and for the frequency effect, the notion of graded activation was introduced into the access phase. That is, to explain the perceptual effects associated with high-frequency words, it was proposed that such words are more highly activated than their low-frequency competitors, and that this results in their faster retrieval (1987:93). The activation mechanism also admits into the cohort lexical items which, though they match imperfectly with the input, are nevertheless the word which the speaker intends should be recognised. Marslen-Wilson (1987:95) notes the problems of variability discussed above, and accepts that hearers can recover from mispronunciations, whether deliberate or inadvertent:

> To accommodate this type of result, the model must find some way of permitting deviant words to enter the cohort.

The way in which the activation mechanism can solve this problem is outlined as follows:

> [Inclusion in the cohort] is not a decision process which requires all-or-none matching, since to discriminate the correct candidate it is not necessary to systematically reduce the cohort to a single member. Selection does not depend on

simple presence or absence in the cohort, but on relative goodness of fit to the sensory input. This makes it in principle possible for candidates that do not fully match the sensory input to participate nonetheless in the recognition process.

Thus, although the hearer presented with the form *shigarette* will presumably still access word candidates with the initial *shi* syllable, the cohort will also contain words which match the input less well; the distance between the word candidate and the sensory input will be reflected in the level of activation of the former. Thus, at an early stage of processing of the form *shigarette*, candidates such as *shilling*, *shiver* and *chicane* will be highly activated on the basis of their close match with the input, while *cigarette*, though present in the cohort, will be activated to a limited degree. At the retrieval phase, however, the activation levels of *shilling*, *shiver* and *chicane* will decrease because of their failure to match closely with the continuing acoustic-phonetic specification of the input (and perhaps also because they fail to meet contextual requirements) while, conversely, *cigarette*'s level of activation will be maintained or even boosted. At some point the differential between *cigarette*'s level of activation and that of the other candidates will reach a criterial level, and *cigarette* will be recognised.

The activation aspect of the later version of the Cohort Model means that it bears certain similarities to the TRACE model (McClelland and Elman 1986). This model consists of a large number of units, organised into three levels and corresponding to phonetic features (such as 'acuteness', 'diffuseness' and 'vocalic'), phonemes and words. Each of these units may be thought of as a hypothesis that the object it represents has occurred at a particular point in time relative to the onset of the utterance. When input is received which supports a particular hypothesis at, say, the featural level, the relevant featural unit and the related units at the other two levels are activated, but other units on the same level are inhibited. This lateral inhibition component is one feature which differentiates TRACE and the 1987 version of Cohort. Bard (forthcoming) has pointed out that the lateral inhibition mechanism may not be necessary to account for results reported to date in the literature.

Context plays a rather different role in the new model, compared with the original version. The top-down and bottom-up knowledge sources do not interact:

> ... no top-down interactions of any sort are permitted. Different types of information are integrated together on-line to produce the perceptual output of the system, but they do not interact in the conventional sense. (Marslen-Wilson 1987:97).

Rather than operating at the level of retrieval, the contextual cues are utilised at the integration stage, when lexical candidates are assessed against the developing syntactic-semantic framework. This results in a model which is serial and autonomous, properties which are considered to be desirable in a theory (Crain and Steedman 1985). To capture the phenomenon of early recognition, Marslen-Wilson suggests that there is

> ... a form of on-line competition between the most salient candidates (those most strongly activated by the sensory input) to occupy the available sites in the higher-level representation. Once the appropriate senses associated with a given word-form have been bound to these locations in the representation, then we can say that recognition has taken place. (Marslen-Wilson 1987:98).

One of the weaknesses of the new formulation of the cohort model is that the criterial level of activation differential which results in recognition is never fully specified (13). Thus, while the earlier version of the theory enabled very precise predictions to be made about the point at which a word would be recognised, the 1987 model is not amenable to empirical evaluation to the same extent. A second flaw in the new model is that the nature of the featural analysis in which the hearer engages is under-specified; depending on the feature matrix which is selected, a wide variety of phonological forms might be permitted into the cohort activated by a given input. In fact, studies such as that of Miller and Nicely (1955) indicate that types of perceptual confusions which hearers permit are highly constrained: when hearers were asked to identify CV sequences in noise, they found that voicing and nasality were highly distinctive features, while, at the other extreme, place of articulation tended to be difficult to identify. Permitting candidates to enter the cohort when they match the input imperfectly at the phonemic level clearly has a number of advantages; yet, interestingly, estimates of cohort structure which form the basis of experimental analysis still tend to be phoneme-based (see, for example, Marslen-Wilson 1988).

How adequately, then, does the Cohort Model account for the findings presented in Sections 2 and 3 of this chapter? By allowing top-down information to influence the retrieval process, the theory can handle the context effects reported in the literature, as long as these effects are confined to the retrieval phase. However, if it could be shown that context effects were present during access (an issue which is still disputed, as we have seen), the model would not be able to account for this finding. Although the predictions of the 1987 version with respect to use of context are less clear-cut, it ought still to be possible to identify an initial phase in the processing of words in which any apparent effects of context are attributable to random factors, since it is claimed that the first 150 msec approximately are devoted to assembling the word-initial cohort.

The variability problem is less adequately handled by the earlier version of the theory. The inability of the 1980 model to account for the recognition of heavily modified words has by now been widely discussed in the literature. The 1987 version attempts to mend this deficiency by advancing the feature of activation, and this property certainly seems, in principle, to allow for the fact that some tokens may be more intelligible and that some parts of the input may be more informative or more readily processed than others.

The lack of segmentation cues in connected speech remains a problem for both versions of the Cohort model, with their heavy emphasis on the importance of word onsets. If the access phase depends on the hearer's ability to identify the beginnings of words, it is crucial that some means of determining the location of word onsets in the speech stream, pre-lexically, should be available. In the absence of consistent acoustic cues in all but the most carefully articulated speech, the hearer's only other hope of performing this feat seems to reside in recognising the preceding word without appreciable delay. If this mechanism fails in a substantial proportion of cases, the possibility of maintaining word-by-word processing of sentences will be greatly diminished. The evidence of studies such as Pollack and Pickett (1963) and Bard et al. (1988) suggests that this is indeed the case.

Before turning to a description of other models of auditory word recognition, it will be useful to identify two aspects of the 1980 version of the Cohort Model which enable it to be distinguished from other theories which will be discussed later. First, the model makes precise claims about the use of context -- specifically, that its use is confined to the retrieval stage, after the initial 150 msec approximately have been heard; second, it emphasises the importance of the beginnings of words, rather than any other part, as a means of making contact with the stored representations of words in the mental lexicon. On both of these points the model is in conflict with the claims of other theories.

The use of context is one of the major parameters according to which models of auditory (and for that matter, visual) word recognition may be characterised. It is usual to make a distinction between interactive models, like the 1980 version of Cohort, which allow contextual or top-down information to play a part in the identification of words concurrently with acoustic-phonetic or bottom-up information, and serial autonomous models, in which the levels of processing are strictly, and sequentially, ordered. Perhaps the best-known example of the latter type of theory is the Search model proposed by Forster (1976; 1979). This theory is claimed to be able to account for the accessing of words in both the auditory and visual modality, and for both perception and production. It must be said, however, that its experimental basis is located largely in the literature on the visual recognition of words presented in isolation; indeed, the visual recognition process is the one which is most fully formulated in this theory. However, it is the applicability of the model to auditory word recognition which is at issue here.

Forster argues that, since essentially the same information about words is required for any language processing activity, be it language production or perception, reading or writing, the use of several different lexicons is unparsimonious. He suggests that the full specification of words is held in a single location, and that access to this master file is gained through one of several peripheral devices which are attuned to the requirements of the modality in question. Thus, the reader will find the specification of a word via the orthographic access file, while the hearer will use the phonological access file, and the speaker the semantic/syntactic access file. Each access file is segregated into 'bins' which are organised according to some criterion relevant to the modality in question; while alphabetic ordering may be appropriate to the orthographic access file, Forster is non-committal about the principles which govern the organisation of the phonological file. Within each bin, entries are organised according to frequency, with the most common words at the beginning of the bin and the least common at the end. When a word is heard, it is subjected to some (unspecified) processing which enables the hearer to locate the appropriate bin in the phonological access file; the entry in this file includes a pointer which records the address of the word's full specification in the master file. The bin is searched sequentially, from the entry with the highest frequency score to the entry with the lowest. The theory is thus able to account for the word frequency effect.

However, the Search model is less successful in its handling of other findings. First, its specification of the phonological mapping process whereby entry is gained to the phonological access file is extremely vague, and fails to address the issues of segmentation, variability and post-offset access. Second, it does not permit context effects. Unlike the phonological mapping issue, this deficiency cannot be explained in terms of a mere oversight, since Forster argues strongly against the validity of the findings relating to the context effect, claiming that they are a function not of the access process but of the response mechanism. While the model can account for contextual effects arising from associated words (by postulating inter-word connections in the master file), it cannot account successfully for the effects of sentential context which have often been reported.

Even if the claims of serial autonomous models with respect to contextual effects could be disproved, interactive theories are still allowed a considerable degree of latitude in their specification of the role of context. For example, while the Cohort Model restricts the use of context to a specific, post-access phase, the Logogen Model (Morton 1969; 1979; 1982) permits the use of contextual information at any stage during the recognition of a word. In Morton's view, the mental lexicon consists of an array of word recognition elements or logogens, one for each word known to the hearer. A threshold is associated with each logogen in the system; when sufficient information of any kind accumulates, the threshold is reached, the logogen fires, and recognition occurs.

Consider first the process of the recognition of isolated words. When a word is heard (or indeed seen), the hearer engages in some preliminary processing of unspecified nature; this results in all of the logogens which share some formal feature with the input being activated to some degree. Thus, when the hearer encounters the word *cat*, the logogens for *cut*, *caught* and *cot*, which share the same consonants as the stimulus, will be activated, though presumably to a lesser extent than the logogen for *cat* itself (14). All logogens have associated with them some pre-set threshold value. When the logogen for a particular word is activated to the degree that the threshold is reached, the logogen fires and that word is recognised. Morton suggests that words of high frequency have lower thresholds than words of low frequency; this results in the former being recognised more readily than the latter.

In the Logogen Model, context effects are accounted for by the interaction between the logogen system itself and the cognitive system, which is responsible for developing, word by word as the input is heard, a high-level interpretation of the sentence. When a sentence fragment is heard, the cognitive system primes all words which are consistent with that sentence fragment, so that less activation is needed to reach the threshold value.

The main argument against this kind of strongly interactive model is that it is in principle untestable because the usual procedure, of holding constant all elements but one in order to test this component, cannot be followed (Simon 1969; Crain and Steedman 1985). A type of interaction which does not possess this undesirable property is weak interaction: higher level components can be called upon to determine which existing hypotheses should be discontinued, but the higher-level components cannot themselves determine which hypotheses should be initiated. This, of course, is the state of affairs with the earlier version of the Cohort Model. Presumably it would be possible to claim that a weakly interactive system could account for results which apparently indicated contextual effects from the very outset of processing of a word, by making the delay to introduction of top-down processing infinitesimally small, but it would seem that this proposition is itself as untestable as the strong interaction account.

It is clear, then, that the Cohort Model presents only one possible view of the role of context in auditory word recognition. We will turn now to a discussion of the second aspect of the model which sets it apart from other models, namely, the manner in which it proposes that phonological mapping between input and lexicon is achieved. As we have seen, there are important differences between the earlier and later versions of the model with respect to this issue. While the first version proposes a strictly phonemic match between the input and the words which are accessed, the second version recognises the need for a more flexible approach which can admit into the word initial cohort lexical items which, though they appear to match rather imperfectly with the input stimulus, are in fact closer to the phonological specification of the word intended by the speaker. Yet both versions of the theory are in agreement over the suggestion that it is the initial portion of words, rather than any other, which enables the speaker to make contact with the mental lexicon. Ideally, if this proposal is to be viable, the hearer should have access to unambiguous cues in the speech stream as to the locations of the onsets of words. Yet, as we saw in Section 2, the speech signal provides few such cues, and those which can be identified are not present consistently even in moderately casual speech. The alternative strategy appears to be for the hearer to identify the location of a word's onset by the efficient processing of the word which precedes it; a corollary of the identification of the end of one word is the pinpointing of the beginning of the next. However, it was demonstrated in Section 3 that the a priori requirement of this approach, that words should be recognised by their acoustic offsets, is often not met. A more serious problem is that words are often not recognised in the order in which they are uttered by the speaker.

It was pointed out in sections 2 and 3 that if some part of the word was to be given special status with respect to the process of phonological mapping between the input and the lexicon, lexically stressed syllables were in some sense more attractive candidates than word-initial syllables. Several writers have more or less tentatively proposed models in which stressed syllables, and other syllables containing full vowels, are used to gain access to lexical representations. Members of this family of models have emphasised different aspects of lexically stressed syllables which might be useful in access. On the one hand, lexical stress itself can be seen as a property of entries in the mental lexicon, with stressed syllables in the speech stream mapped on to the appropriate portion (Cutler 1976; Gee and Grosjean 1987); this proposal utilises the fact that syllables realised as stressed have potentially useful acoustic characteristics. On the other hand, some writers have focussed not on lexical stress itself but on the related notion of full vowel quality, which may be found in unstressed as well as stressed syllables (Cutler and Norris 1988).

Among the first to sketch out a model utilising the prosodic aspects of speech was

Cutler (1976:59):

> Consider the hypothesis that the mental lexicon is so arranged that one (or *the*) primary principle by which a word is classified is the nature of its stressed syllable (or the vocalic nucleus of that syllable). When a stressed syllable is identified, then, the sentence processor can begin immediately to locate in the mental lexicon the word of which it is a part, using information not only about the stressed syllable itself but about the number of unstressed syllables immediately before and after it.

Cutler herself clearly no longer espouses this view, attributing the role of prosodic information to a different stage of processing (Cutler 1986; Cutler and Norris 1988), but others have made a similar suggestion. Grosjean and Gee (1987) have presented a model, which they admit is under-specified. They assume that hearers first submit the acoustic signal to some sort of analysis which results in a phonetic string marked for weak and strong (i.e. unstressed and stressed) syllables: this representation mediates between the acoustic signal and the lexicon. The distinction between stressed and unstressed syllables is crucial. According to their model, the process of lexical access proceeds as follows (1987:144)

> On the one hand, stressed syllables (and only they) are used to initiate a lexical search. On the other hand, and concurrently, the weak syllables located on either side of the stressed syllable (functors, affixes, weak syllables of content words) are identified by means of a pattern-recognition-like analysis and with the help of the listener's knowledge of phonotactic and morphophonemic rules.

Grosjean and Gee's notion of the stressed-unstressed distinction is presented in metrical terms (Hayes 1980, 1982; Liberman and Prince 1977; Selkirk 1978, 1980; Kiparsky 1979; McCarthy 1979) in which utterances are considered to consist of weak and strong (in Grosjean and Gee's terms, stressed and unstressed) elements. A prerequisite of the strong version of this model is clearly the ability of listeners to distinguish between stressed and unstressed syllables in the speech stream, prior to lexical access. As the authors point out, there is little direct and non-controversial evidence to support their view: this observation extends to this prerequisite also. A weaker version of the theory might tolerate a large number of errors in identifying stressed syllables in the speech signal, resulting in numerous unsuccessful lexical access attempts (S. Isard, personal communication); Bard (forthcoming) argues that this strategy is unacceptably inefficient. Gee and Grosjean do not distinguish between weak and strong versions of the model; the weak version is both less testable and potentially extremely inefficient.

Plausible though this account may seem, some existing experimental evidence fails to support it. Shillcock (in press) describes a number of experiments using cross-modal priming and word naming to investigate various effects relating to the processing of words, specifically of lexical items embedded in other words (such as *bat* in *wombat*). One of these experiments (which used word naming) concerned the effects of high frequency polysyllabic competitors to low frequency monosyllabic words. In one condition the overlap between the two items was word-initial (e.g. *numb* vs *number*), while in another condition it was word-final (e.g. *sheen* vs. *machine*). Competition from at least one high-frequency polysyllabic competitor sharing the initial syllable was sufficient to prejudice recognition in white noise, but there was not equivalent competition from words sharing word-final specifications. Shillcock argues that this result is inconsistent with Gee and Grosjean's proposed model, which predicts that words like *number* would be accessed via their initial syllables while words like *machine* would be accessed by their final syllable; this in turn appears to predict equal competition in both of the conditions in the experiment reported by Shillcock, contrary to the actual findings of his study.

Grosjean and Gee's model has little to say about the detailed processes involved in word recognition, and is thus difficult to assess with respect to the context issue discussed above. They suggest that functors and other unstressed syllables can be recognised in a right-to-left manner, but do not offer a detailed account of this. Instead, they see their role as drawing to the attention of the psycholinguistic community the need to incorporate prosodic elements into word recognition models. As far as the segmentation and variability problems are concerned, models which propose that lexical access is initiated via stressed syllables have much to commend them, in theory at least. First, it is not necessary for word boundaries to be identified: lexical access is triggered by stressed syllables rather than word beginnings.

While it is true that some studies have pointed to a relationship between stress and acoustic parameters such as amplitude, duration and F_0 movement, the correlation between these features is by no means perfect. Until there is experimental evidence that hearers can unambiguously distinguish stressed and unstressed syllables, the model (at least in its strong version) potentially suffers from deficiencies similar to those which characterise 'word-onset' models such as Cohort.

As far as the variability problem is concerned, however, the stressed syllable models fare rather better. Much of the attraction of adopting the stressed syllable as the phonological key to the mental lexicon resides in their relative immunity to reduction phenomena and other phonological modifications; thus, they preserve a relatively direct relationship between their acoustic manifestations and their underlying phonological form. Thus, stressed syllable models might well offer a successful treatment of the phonological access issue.

The model proposed by Cutler and Norris (1988) concentrates on the vowel quality of syllables rather than on their lexical stress value. They propose that when the hearer encounters a portion of the speech stream with particular acoustic properties, it is hypothesised that a full vowel has been encountered and a tentative word boundary is inserted at the onset of the syllable containing the vowel; lexical access attempts are made on the basis of this input. While this strategy has the advantage of utilising distributional properties of English (Cutler and Carter 1988) and of accessing words by potentially informative parts (Altmann and Carter 1989), the results which it models can be accounted for without recourse to a component specifically designed to identify unreduced vowels as Bard (forthcoming) has pointed out (8).

When the experiments reported in the following chapters were being designed, Cutler and Norris's model had not been proposed. It is therefore omitted from consideration in the following section, which discusses the hypotheses examined in the study, but its predictions will be considered with respect to the experimental results presented in later chapters.

5. Summary and general discussion

The overall aim of this chapter was to explore the extent to which existing models of word recognition accounted for the known facts about hearers' behaviour in recognising spoken words. Table 2.1. summarises the manner in which the models considered have addressed the three issues identified earlier (use of context; phonological access; post-offset recognition).

Although all of the models have a mechanism for dealing with context, the full range of logical possibilities is represented: Logogen allows contextual effects during access, the 1980 version of the Cohort model during retrieval, and Search and the later version of Cohort exclude them from the access and retrieval stages of processing altogether. The stressed syllable model, as proposed by Gee and Grosjean, allows it during retrieval, by postulating functors for unstressed sequences. Clearly, there is room for further investigation of the context issue.

Table 2.1: approaches of word recognition models to three issues in word recognition

	Context	Phonological mapping	Post-offset recognition
Cohort-1	during retrieval	word onset	--
Cohort-2	during integration	word onset	cohort resolution
Search	after retrieval	--	--
Logogen	during access	--	--
Stressed Syllable	--	stressed syllable	during retrieval (unstressed forms)

Key

Cohort-1: Cohort Model as described by Marslen-Wilson and Tyler (1980)
Cohort-2: Cohort Model as described by Marslen-Wilson (1987)
Search: Forster (1976; 1979)
Logogen: Morton (1969; 1979; 1982)
Stressed Syllable: e.g. Cutler (1976), Grosjean and Gee (1987)

The phonological access issue also evokes two dramatically differing approaches: while the Cohort Model bases access on the beginning of the word, the Stressed Syllable Model proposes that the relative invariance of stressed syllables should be exploited for this purpose. Again, further research is necessary in order to distinguish between these alternatives.

The last issue, post-offset recognition, is only addressed by the two models which have been fairly recently proposed, namely the second version of the Cohort Model and the stressed syllable model. While the Cohort Model allows late recognition for homophones (and of partial homophones, such as might occur when short words are embedded in longer words, as is the case with *you/ewe* in *unicorn*), it is claimed that under normal circumstances polysyllabic words heard in context should regularly be recognised by offset. As far as the stressed syllable model is concerned, only weak forms, which it is proposed will usually be functors, are considered as candidates for late recognition.

The context, phonological access and post-offset recognition issues form the basis of the experimental work to be reported in the following chapters. The inter-related nature of these three aspects of word recognition should be noted. It has been argued above that phonological mapping based on word onsets should be difficult for the hearer because of the lack of segmentation cues and the variability inherent in fluent speech, factors which must be partly responsible for the post-offset recognition phenomenon. It might be suggested that hearers compensate for the relative unintelligibility of words by the use of contextual information; yet if a substantial proportion of words in fluent speech fail to be recognised promptly, on what basis can the linguistic context be made available? The prevalent view of word recognition is that it is a process which proceeds word-by-word, in a left-to-right fashion. If such a view is correct, then listeners presumably have available to them, as they process a given word, a syntactic and semantic framework constructed on the basis of previously recognised words. Such a framework would be a valuable asset in the recognition of words. Yet recently, various writers have challenged this word-by-word, left-to-right view (e.g. Grosjean 1985; Bard et al. 1988). Their experimental results indicate that hearers sometimes fail to recognise words until subsequent acoustic material is presented. If an item early in an utterance is unrecognised, it clearly cannot be integrated into the higher-level description of the utterance which is made available for the processing of subsequent words. Thus, to the extent that preceding words in an utterance fail to be recognised, the framework which ought to contribute to efficient processing is impoverished.

Of the five models discussed in this chapter, the one which is most highly specified in terms of the issues under consideration is Cohort-2. However, at the time when the experiments reported here were being designed, the later version of the theory was not yet formulated. Similarly, the stressed syllable model outlined above did not appear in the literature until after this study was conducted, although some model of this kind was assumed in the design of the experiments. The hypotheses which are tested in the subsequent chapters were therefore framed in terms of Cohort-1, which was the most fully specified model available at that time.

As far as phonological access is concerned, the Cohort Model predicts that hearers use the onsets of words to gain entry to the mental lexicon. This *phonological access hypothesis* makes no allowance for word-initial syllables of differing levels of intelligibility, such as might arise, for example, because of differences in stress level or of speech style. Although the later version of the model contains a feature-based mechanism which takes account of the variability of word tokens, it still falls prey to the segmentation problem to the extent that it relies specifically on word onsets. The alternative proposal, that stressed syllables are used to access the lexicon, potentially has a number of advantages. In particular, under conditions in which phonological modifications are highly likely to occur (for example, in conversational speech) stressed syllables are relatively invariant.

With respect to the use of context, the Cohort Model also makes precise predictions, which are at odds with the claims of the other models which address the issue. According to the Cohort Model, lexical access is achieved purely on the basis of phonological mapping, with no mediation of context. That is, this model makes a *context-free hypothesis* about the lexical access phase; if we could examine the word candidates which were accessed when a token was heard in isolation, we should find that they did not differ significantly from those which were accessed when the same token was heard in context.

In their discussion of the structure of the Cohort Model, Marslen-Wilson and Tyler (1980) make no specific reference to the post-offset recognition phenomenon, but it is clear that the majority of words need to be recognised by offset if hearers are to be aware as often as possible of the location of word onsets. In discussing the later version of the model, Marslen-Wilson (1987) has claimed that polysyllabic words, and any words heard in context, should be recognised before their acoustic offsets. This *early* *recognition hypothesis* should be easily testable. On the basis of other studies which have been reported, it is likely that the early recognition phenomenon will be common in carefully articulated, read speech, but perhaps less so in conversational tokens.

While the first two hypotheses -- the phonological access hypothesis and the context-free hypothesis -- relate to the access phase of processing, the early recognition hypothesis is concerned with retrieval. The fact that both phases are are to be examined places important constraints on the selection of an experimental methodology. This question will be addressed in detail in Chapter 3.

Chapter Notes

(1) In fact, Marslen-Wilson (1987) identifies a further process, which he terms 'integration' and which involves the construction of a syntactic-semantic representation of the sentence. But this aspect will not be considered here.

(2) 'Context' should include syntactic, semantic and pragmatic elements. In practice, however, the first two sources of contextual information have been studied at the expense of the third in psycholinguistic studies of word recognition. The research reported here will concentrate on the first two sources of contextual information.

(3) They explain their results in terms of limited lookahead by speakers: for example, when speakers reach a syntactic boundary, their processing capacity is temporarily exhausted and they fail to apprehend the following segment which is potentially the second part of the environment required for the phonological rule to apply. It is hard to see how such an account can explain the word frequency effect, however. It is tempting to suggest that the explanation lies in some form of co-operative strategy by speakers, who wish to structure their message in such a way as to assist hearers at points in the sentence where they are likely to encounter difficulty; however, one must be cautious of attributing to speakers the ability to carry through such good intentions (see, for example, Bard and Anderson 1982; Shockey and Bond 1980).

(4) The same terminology has been used by some writers (e.g. Cutler 1986; Cutler and Norris 1988) to refer to vowel quality and other aspects of the pronunciations of words.

(5) It is beyond the scope of this study to attempt to distinguish between the shades of theoretical meaning attached to the many and varied approaches to prosody. See Cutler and Ladd (1983) for a discussion.

(6) These techniques will be considered in more detail in Chapter 3, which will examine methodological issues. The experimental tasks are outlined here for the benefit of the reader unfamiliar with them, since knowledge of them will be assumed in the following sections.

(7) Note that adoption of these acoustic-phonetic features as the means of identifying strong syllables would tend to succeed in identifying only a subset of the full vowels, probably those which are stressed and monophthongal. Stressed and unstressed diphthongs are characterised by movement of the formants, while the duration of phonologically short, monophthongal, unstressed vowels is likely to be close to that of schwa. However, it is not possible to cite experimental evidence on this point; Harrington (personal communication) has pointed out that virtually no research has been done on the duration of monophthongal versus other unstressed vowels, and that part of the reason for this is the difficulty of finding minimal pairs which could be examined.

(8) Bard (in preparation) has questioned the need for postulating a strong syllable segmentation strategy. She argues that results such as those of Cutler and Norris (1988) are consistent with a model of auditory word recognition in which word hypotheses are generated with differing levels of activation: some portions of the speech signal (e.g. strong syllables) activate fewer candidates than others because of their greater ability to partition the lexicon. Such candidates receive a high level of activation and inhibit the level of activation of competitors. When candidates overlap, as in Cutler and Norris' materials, they either inhibit each other, and are consequently recognised comparatively slowly, or else recognition of the embedded word is delayed until the activation of the overlapping hypotheses has died away.

(9) In a recent paper, Blutner and Sommer (1988) have used the cross-modal priming technique to investigate further the relationship between word recognition and focus. They found that they could replicate Swinney's 1979 result for ambiguous words only if the items in question were focussed. That is, if words were in focussed position, both senses would be available at offset, but only the contextually appropriate sense would be available after a short delay. For unfocussed words, the same result was found when probes were delayed, but probes at offset showed no evidence of the activation of the inappropriate sense.

(10) One of the most widely studied factors in word recognition, apart from those considered in this chapter, is word frequency. The word frequency effect is well-known in visual recognition studies (e.g. Rubenstein, Garfield and Millikan 1970; Forster and Chambers 1973). Evidence has also accumulated indicating a frequency effect in the auditory modality. Several studies which used the listening-in-noise paradigm (e.g. Broadbent 1967) showed that words of high frequency were more intelligible than words of low frequency. Grosjean (1980), using the gating paradigm, found that high-frequency words could be identified on the basis of less acoustic-phonetic input than low-frequency words. Marslen-Wilson (1987) reports an experiment which used a lexical decision task to compare reaction times to high and low-frequency words matched for uniqueness point (e.g. *street/streak*); measuring the reaction time from the uniqueness point rather than the word onset, he found a considerable advantage for high-frequency words.

Despite the consistency of these results, there remained until recently some doubt as to whether the frequency effect was genuinely perceptual (in which case it must be incorporated into models of word recognition) or post-perceptual. Marslen-Wilson et al. (1987) have attempted to resolve this issue using a cross-modal priming task. They selected pairs of words which had similar initial portions, e.g. *captain* and *captive*, but which differed with respect to word frequency (the former being more common than the latter). They placed visual probes, to which subjects made lexical decisions, in two positions: 'early', just before the uniqueness point, and 'late', at the end of the word. The visual probe for the word *captain* was *ship* and for *captive* was *guard*; in the early probe condition, there was more facilitation of *ship* than of *guard*, irrespective of the word which was auditorily presented, but this advantage had disappeared by the late probe position. Since the frequency effect is found only at the early probe position, before the auditory word can have been recognised, the result must reflect the recognition process itself, and not some post-perceptual bias. Thus, it appears that models of word recognition need to incorporate a mechanism for preferential treatment of high-frequency words. However, the results of lexical decision tasks (which form part of the cross-modal priming technique) might be open to criticism, since Balota and Chumbley (1984) have argued that that this task involves processes which have little to do with lexical access but which might result in a word frequency effect.

(11) In discussing the role of syntax in speech processing in the light of these results, they suggest (1983:418) that syntactic constraints 'function to reduce the amount of sensory input needed for recognition of ... suffixes'. While it is true that inflectional suffixes at least are often poorly represented in the acoustic signal (Gimson 1980:295), it is not clear why hearers would wish to recognise suffixes, unless it is to develop a syntactic analysis of the sentence. Additionally, such a strategy, adopted for whatever reason, would be fruitless in many cases since the majority of words in an English sentence are not morphologically marked for word class. Tyler and Marslen-Wilson (1986) examined gating responses to polymorphemic words and found that under conditions of strong syntactic constraint, the entire word, complete with inflection, was correctly identified as soon as the base was identified, but under weak syntactic constraints, identification of the inflected form fell on average 200 msec later than the stem identification point. The effect of semantic constraint was to allow earlier identification of the base form without changing the identification point of the inflected form.

(12) In fact, it is probable that a large number of word *tokens* remain homophonous, or ambiguous, beyond their offsets. Apart from the obvious case of functors (e.g. [@ z] could be the phonetic representation of *has, is, as* or *us*), short content words can also display this ambiguity of phonetic form because of the deletion and assimilation processes which frequently operate. Consider, for example, the sequence [h a m b a g]; the portion [h a m] could be interpreted as either *hand* or *ham*, but in certain sentence structures only right context could signal the correct interpretation.

(13) Although this criticism also applied to the TRACE model as described in this chapter, recent attempts have been made to remedy the deficiency (McClelland and Rumelhart 1988).

(14) This account assumes that the token which is heard approximates more closely to *cat* than to any other word -- an assumption which the evidence of Section 2 has suggested may be questionable.

CHAPTER THREE

Methodology

1. Introduction

In this chapter the use of the experimental paradigm employed in the study is justified. The requirements of the study are outlined in section 2 and the extent to which the existing paradigms meet these requirements is discussed in section 3. In section 4, gating, the paradigm used in the research, is examined in more detail: aspects of methodology used in earlier gating studies are described (section 4.1); the appropriateness of the gating paradigm for the present research is assessed and its limitations discussed (section 4.2).

2. Requirements of the study

The choice of experimental paradigm in this study is constrained by the nature of the hypotheses detailed in Chapter 2. While the early recognition hypothesis related to the retrieval stage of processing, the other two issues, the context-free and phonological access hypotheses, concerned the characteristics of the word candidates considered during lexical access. It is thus necessary to identify an experimental technique which can shed light on both of these processes. As the following section will show, the technique which best fulfills these requirements is the gating paradigm (Pollack and Pickett 1963; Pickett and Pollack 1963; Grosjean 1980; Tyler and Wessels 1983, 1984; Cotton and Grosjean 1984; Tyler 1984; Salasoo and Pisoni 1985; Grosjean 1985; Bard et al.,1988).

3. Existing paradigms

Those paradigms commonly used in word recognition studies are summarised in Table 3.1., which shows, for each experimental technique, the dependent variables which can be measured.

Table 3.1. Experimental techniques for examining word recognition

DEPENDENT VARIABLES	Temporal Point of Access/Recognition	Error Patterns (quantitative)	Error Patterns (qualitative)
PARADIGM			
1. phoneme monitoring	*		
2. syllable monitoring	*		
3. word monitoring	*		
4. monitoring for mispronunciations	*		*
5. lexical decision	*		*
6. cross-modal priming	*		
7. shadowing	*	*	*
8. word identification	*	*	*
9. gating	*	*	*

The characteristic feature of the first seven paradigms (phoneme, syllable and word monitoring; monitoring for mispronunciations; lexical decision; cross-modal priming; shadowing) is their use of reaction time as a dependent variable; these tasks are collectively termed 'on-line', and are thought to reflect normal word-recognition processes since they do not appeal to explicit introspection on the part of the subject. The common assumption of such studies is that the time between the onset of the test stimulus and the initiation of the subject's response corresponds to the time taken to access the stimulus in the mental lexicon (1).

In the first four paradigms, subjects are required to press a key when they hear the target phoneme (Foss, 1969), syllable (Savin and Bever 1970) or word (Marslen-Wilson and Tyler (1980)); the dependent variable is the speed with which they respond. The monitoring-for-mispronunciations paradigm (Cole 1973) is a variant in which the target is not prespecified; instead, subjects respond to any word which they realise has been mispronounced. In addition to reaction time, the number of mispronunciations detected may also be measured.

The lexical decision task has been employed in both the auditory (Cutler and Clifton 1984) and visual modalities (Rubenstein et al. 1970). Subjects are presented with lists of stimuli and must judge whether each item is a word or a non-word; their speed of response is the dependent variable. Visual lexical decision is a sub-task in the cross-modal priming paradigm (Swinney 1979). Subjects hear sentences containing test words. At the offset of the test word the visual stimulus, either a word or a non-word, is presented; response time has been found to vary with the degree of relatedness between the visual and auditory stimuli.

The shadowing task has been employed in several studies by Marslen-Wilson (1973,1975) and others (Cherry 1953; Moray 1959). Subjects listen to tapes over headphones and attempt to repeat what they hear as soon as possible after hearing it, monitoring their own performance while processing the incoming signal. Reaction times are measured from the onset of an input word to the onset of its output counterpart. The accuracy of the shadowing performance forms a second dependent variable. Some qualitative analysis of the word candidates accessed can be conducted when subjects produce errors, but this is not an efficient or reliable method of collecting such data.

In contrast to the first seven paradigms, the last two (word identification; gating) typically impose no temporal constraint on the subject's response other than that introduced by the arrival of the following stimulus. This characteristic has, in fact, been one of the major sources of objection to the use of such tasks (this issue is discussed in more detail in section 2.1). The word identification paradigm has a long tradition in word recognition research, going back at least as far as Bagley (1900). Variants of the paradigm which involved presenting stimuli in noise were common in the 1950s (e.g. Miller, Heise and Lichten 1951). Although the standard dependent variable in such studies is the number of words correctly identified, speed of recognition can also be measured (Nooteboom 1981), and some qualitative analysis is possible. However, if the issue which is of interest relates to words accessed by hearers on the basis of some specific portion of the utterance (for example, the initial syllable) the identification paradigm is unsuitable because subjects will have heard more than just the critical portion of the stimulus before making their response.

The gating task was pioneered by Grosjean (1980). In this paradigm subjects hear successively longer portions of a stimulus, and are required to respond, at each pass, with a response at the identity of the target and possibly also with a rating of their confidence in the accuracy of the response. Grosjean likens the technique to that used by Pollack and Pickett (1963), who presented sentences to subjects word by word, such that on the first pass, only the first word of the sentence was heard; on the second, the first two words; and so on until all of the sentence had been presented. The paradigm proposed by Grosjean is also characterised by the systematic lengthening, by small degrees, of the portion of the stimulus heard by subjects; but, typically, it is individual words which are gated, at increments of some tens of milliseconds. Though the point at which lexical access occurs can only be measured in rather crude terms, being limited by the size of incremental unit selected, the advantage of the paradigm lies in the record of responses made by subjects prior to correct identification of the target word, which has been assumed by all the researchers who have used the technique to reflect the candidate set activated in the mental lexicon. This aspect of the data collected in gating experiments gives scope for a further set of dependent variables, such as the number of candidate types proposed at each gate (Grosjean 1980), the characteristics of the candidates (Tyler 1984) and the contextual appropriateness of the responses (Tyler and Wessels 1983).

The hypotheses discussed in Chapter 2 concerned, on the one hand, the top-down and bottom-up appropriateness of the candidates accessed by hearers when a particular stimulus is presented, and on the other, the point at which words are recognised or retrieved. Any of the techniques discussed in this section can shed light on the latter issue. But if one is to examine the characteristics of those words which are not the target but are nonetheless accessed by the hearer (in terms of Table 3.1., to conduct a qualitative analysis of the errors), only a subset of the available techniques are suitable. Of these, the one which provides the richest source of data is gating. This paradigm will therefore be used in the main experiments of this study.

In the following section, the gating paradigm is reviewed in more detail; those studies which have used it are summarised, and its advantages and disadvantages are considered.

4. The Gating Paradigm

Although the gating paradigm has been used to investigate several linguistic phenomena, such as sign language (Grosjean and Lane 1981) and the processing of prosodic information (Grosjean 1983) the discussion in this section will be confined to the use of the technique to study spoken word recognition. The aim of this section is merely a discussion of methodological aspects of the paradigm, rather than a summary of experimental findings; the theoretical implications of the studies were discussed in Chapter 2.

4.1. Summary of Earlier Studies using the Gating Paradigm

Methodological aspects of the studies discussed in this chapter are summarised in Table 3.2.

The earliest gating studies were by Pickett and Pollack (1963) and Pollack and Pickett (1963). These two studies differ from several of the others described in this section in that the incremental unit selected was the word, rather than some arbitrary sub-word-sized portion. They are included here, however, for historical interest, and also because the second study is one of the first which investigated conversational speech.

Table 3.2. Summary of methodological aspects of gating studies.

VARIABLES	incremental unit	presentation type	speech type	response type	pressure to respond	gating direction	number of words gated per sentence*
STUDY							
Pickett and Pollack 1963	word	successive	read	written	untimed	forward	-
Pollack and Pickett 1963	word	successive	spontaneous	written	untimed	forward	-
Grosjean 1980	30ms	successive	read	written	untimed	forward	1
Cotton and Grosjean 1984	30ms	individual	read	written	untimed	forward	1
Tyler and Wessels 1983	50ms	successive	read	written	untimed	forward	1
Tyler and Wessels 1984	50ms	individual	read	spoken	timed	forward	1

* within-word gating studies only

Table 3.2. (continued): **Summary of methodological aspects of gating studies.**

VARIABLES	incremental unit	presentation type	speech type	response type	pressure to respond	gating direction	number of words gated per sentence*
STUDY							
Tyler 1984	50ms	successive	read	written	untimed	forward	1
Salasoo and Pisoni 1985 (experiment 1)	50ms	successive	read	written	untimed	forward and backward	4-5
Salasoo and Pisoni 1985 (experiment 2)	50ms	individual	read	written	untimed	forward and backward	4-5
Grosjean 1985	50ms	successive	read	written	untimed	forward	4
Bard et al 1988	word	successive	spontaneous	written	untimed	forward	-

* within-word gating studies only

The object of investigation in the first paper was the effect of variations in speech rate on the perception of fluent utterances. Four male speakers were recorded reading four (different) paragraphs at three different speaking rates; between six and eight sections from the speech of each talker were selected and gated so that they could be presented to subjects one word at a time, incrementally. Subjects were to write down, after each sentence fragment, what they thought they had heard.

The experiment was replicated under two other conditions, firstly, in the presence of masking noise, and secondly, with short tokens synthetically stretched (in order to determine whether the observed low intelligibility of such tokens was due to the inability of subjects to perceive speech at such short durations: this did not appear to be the case). Due to a methodological shortcoming of this first study, whereby a single subject heard all three versions of the same sequence (always in the order fast rate, followed by medium, followed by slow), the rate effect is impossible to evaluate accurately. An interesting finding, however, was that intelligibility of an item increased with sequence duration, that is, words were more likely to be recognised in longer than in shorter sequences. The intelligibility of earlier words rose as the the sequence acquired additional words.

In the second study, Pollack and Pickett (1963) were still interested in the effects of rate on intelligibility. Rather than coaching speakers to perform at different rates, as they had done in the earlier experiments, they decided to take advantage of the variable nature of speaking rate in spontaneous conversation (2). They surreptitiously recorded the conversation of four female speakers and located sections of speech of approximately 2 seconds' duration. The number of samples varied from speaker to speaker, but was approximately eleven; the number of words in each sample ranged from four to fifteen. When these sequences were presented to subjects for identification in the same manner as in the previous experiment, a similar tendency for sequences containing more words to be more intelligible was found.

Within-word gating was introduced by Grosjean (1980). Test words were presented in 30 millisecond increments to subjects who, at each pass, were to write down what they thought the test word was and to indicate their confidence in the response by marking a horizontal scale which ran from 'very sure' on the left to 'very unsure' on the right. Stimuli were heard in three context conditions: for the test word *camel*, heard in the long-context condition, the first three presentations would be as follows:

1. *At the zoo, the kids rode on the* + first 30 msec of *camel*
2. *At the zoo, the kids rode on the* + first 60 msec of *camel*
3. *At the zoo, the kids rode on the* + first 90 msec of *camel*

The token used in the long-context condition was copied and edited to produce two further conditions: a short-context condition, the first three gates of which were

1. *The kids rode on the* + first 30 msec of *camel*
2. *The kids rode on the* + first 60 msec of *camel*
3. *The kids rode on the* + first 90 msec of *camel*

and a no-context condition:

1. first 30 msec of *camel*
2. first 60 msec of *camel*
3. first 90 msec of *camel*

One of the aims of this study was to validate the paradigm as a technique for the investigation of word recognition. To do this, Grosjean manipulated word frequency, word length, and context, since other accepted techniques had consistently revealed effects associated with these three variables. If the gating paradigm replicated these three results, then the technique ought to be reliable. The results supported Grosjean's prediction that shorter, more common words in a restrictive context would be identified more quickly than longer, rarer words presented in isolation, and would elicit a narrower range of wrong responses: such findings were compatible with those observed using other techniques.

The dependent variables used in the 1980 study reflect the richness of the data elicited by the technique:

(1) the point at which subjects offered the target word as the response and did not subsequently change their minds. This measure, termed the isolation point, could be absolute (number of gates or milliseconds) or relative (percentage of the length of the word);

(2) subjects' confidence rating at the isolation point and at the final presentation (the point at which subjects' confidence rating reached a criterial level was later termed the recognition point);

(3) number of words which at least one subject failed to isolate by the final presentation;

(4) the point at which the target word is offered as a response by at least one subject;

(5) word candidate types offered at the first gate;

(6) word candidate types offered across all gates;

(7) phonotactic composition of the wrong responses;

(8) word frequencies (Kucera and Francis 1967) of the wrong responses;

(9) contextual appropriateness of wrong responses.

One of the issues which must be addressed in gating studies is the treatment of those instances in which subjects fail to recognise a word by its offset. Grosjean's solution, in which he has been followed by most other researchers, was to assign to such tokens an isolation point score equivalent to the duration of the word. Fortunately, in the 1980 study the problem was of manageable proportions, since relatively few words were unrecognised by their final presentation. But if the technique were to be used with less intelligible speech, another solution might need to be adopted.

Grosjean (1980) did not deal with a possible objection to the paradigm, namely that the repeated presentations of portions of the same stimulus to the same subject might enable him or her to narrow down the range of word candidates in a way foreign to normal word recognition processes. Furthermore, subjects might be influenced in their confidence ratings by the knowledge that they would later be hearing the stimulus in full. These objections were met by Cotton and Grosjean (1984), who took a subset of the materials from Grosjean's 1980 study and presented individual (rather than successive) gates to different groups of subjects. In other words, each subject heard only one version of each word, corresponding to a single presentation in Grosjean's 1980 experiment. The materials which were selected for the 1984 study were as follows: thirty-one of the original 48 words from the no-context condition, presented at 30, 120, 210, 300 and 390 msec gate lengths; and forty-six of the original forty-eight short-context stimuli, heard at 30, 90, 150, 210 and 270 msec. The results showed that the successive presentation format did not differ significantly from the individual presentation format in terms of isolation points or error patterns.

Salasoo and Pisoni (1985) also validated the presentation aspect of the technique as one of several modifications to the original paradigm. The aims of their paper were first, to explore the interaction between different sources of knowledge in word recognition, and second, to determine the contributions of word-initial and word-final information. They examined the first variable by presenting (unrelated) meaningful and semantically anomalous sentences, and the second by introducing the innovation of right-to-left gating, which they compared with the more conventional left-to-right kind. In earlier studies using the paradigm, only a single word - usually the last in the sequence presented - was gated; a further modification introduced by Salasoo and Pisoni was the simultaneous gating of *all* the content words in the sentence. This innovation was not productive, since it made the ensuing results extremely difficult to interpret. This multiple gating necessitated a final extension of the original technique: because stimuli were, in the majority of cases, followed by other words in the same sentence, envelope-shaped white noise (rather than silence, as in previous studies) was substituted for the part of the stimulus which had been removed, in order to preserve prosodic cues while removing segmental information. Their efforts to validate the successive presentation format were concentrated in the second of two experiments. For all sixteen sentences (eight meaningful and eight anomalous) used in the first experiment, which had used the successive format, they created eight gated stimuli, ranging in length of test word from 0 msec to the full duration of the word. Although no presentation effect was found for meaningful sentences, subjects could more quickly identify words in anomalous sentences if they were heard in the individual than in the successive presentation format. The reason for this discrepancy cannot be determined unequivocally, since different words and different syntactic structures were used in the two semantic conditions; the authors' own interpretation (1986:225) is that in the semantically anomalous condition

> conflicting top-down semantic and syntactic information may have accumulated with successive repetitions of the same anomalous context to prevent efficient and rapid use of word-initial acoustic-phonetic information.

Tyler and Wessels (1983) used the paradigm to quantify the relative contributions of syntactic and semantic context to the word recognition process, by varying the semantic and syntactic characteristics of the sentence fragment preceding the test word. In view of these aims, the possibility of generating several experimental stimuli from a single token was not open to Tyler and Wessels; rather, they spliced a token of the test word (excised from a neutral carrier sentence) on to four stimuli in which they had varied the amount of semantic and syntactic information available. In a fifth condition, subjects heard the excised word on its own, as a control to assess the contribution of the acoustic information available in the experimental conditions. Subjects again responded at each pass by writing down their response and by indicating their confidence, this time by an explicit numerical measure ranging between one (completely unsure) and ten (absolutely certain). The inclusion of many filler items in this experiment resulted in extremely long experimental sessions: subjects took part in two sessions, each lasting one and three quarter hours. Dependent variables examined in this study were isolation points and recognition points (defined as the gate at which subjects achieved an 80% confidence rating, which they subsequently maintained). In addition, Tyler and Wessels undertook a detailed examination of the record of wrong responses prior to successful recognition, which they discussed in the light of the predictions of the Cohort model.

Tyler and Wessels' 1984 study used a subset of the materials from their 1983 paper. Their aim in this second paper was to overcome a further objection to the paradigm, namely that because subjects in previous gating experiments had been allowed to produce their responses without time pressure, the results reflected not unconscious language processing but rather the outcome of their conscious introspection. They therefore presented subjects with those materials from the earlier experiment whose contexts had been found to be intelligible (i.e. those in which the Dutch function word *te*, which provided the strong syntactic constraint, was correctly perceived by the majority of hearers in a post-test reported in the 1983 paper). Given the emphasis on speed of response, an individual presentation format was used, since subjects' later RTs would no doubt be influenced by having heard parts of the stimulus. In order to limit the size of the experiment, therefore, only five gate sizes were selected: the isolation point (determined from the earlier experiment), the isolation point minus 100 msec, the isolation point minus 50 msec, the isolation point plus 50 msec and the isolation point plus 100 msec. In this 'timed' study, subjects were to respond as in the previous, 'untimed' experiment, except that they were to report their response aloud. When the response latencies were measured (from the offset of the fragment to the onset of the subject's response) they were found to fall within the range observed using other tasks usually considered to be on-line (auditory and visual lexical decision; mispronunciation detection; phoneme monitoring), suggesting that the timed version of the experiment was indeed comparable to on-line tasks. In order to validate the untimed version in a similar manner, isolation points, recognition points and response patterns were compared across the two studies. The relationship between the five conditions examined in the earlier experiment remained the same for all these measures. The authors concluded that both the timed and untimed versions of the task reflect on-line processes, but that since the untimed version is less time-consuming to administer, it is preferable to use it rather than the timed task. To evaluate this point it is worthwhile to consider the elements of the design of the timed study which differentiate it from the untimed version: firstly, the use of the individual presentation format, and secondly, the verbal rather than written responses of the subjects. Both aspects of the design make the timed task laborious from the point of view of the experimenter; as far as the subject is concerned, however, a version in which responses are verbal must be preferable in terms of the length of experimental session required.

Tyler (1984) undertook the most detailed analysis to date of the pattern of wrong responses prior to correct identification of the target word. She examined the data from the experiment reported by Tyler and Wessels (1983) to determine to what extent the predictions of the Cohort model were fulfilled, evaluating the data (the 'elicited cohort') in terms of cohort size, word frequency, number of syllables and syntactic and semantic appropriateness. She was particularly interested in discovering, firstly, what factors were implicated in the assembling of the word-initial cohort, and, secondly, what factors caused this set to be narrowed down to a single member, at which stage recognition is assumed to have occurred.

In a recent paper, Grosjean (1985) extended his original paradigm to investigate the recognition of words after their acoustic offset, a phenomenon which had been represented in his (1980) data as well as in those of Pickett and Pollack (1963) and

Pollack and Pickett (1963). In this experiment, Grosjean gated not only the test word but also the three words following it in the same sentence. By selecting words beginning with stop or fricative consonants, Grosjean was able to locate the onset of his test stimuli with some degree of consistency, but the preposition, determiner and noun which followed were not amenable to similar treatment. Grosjean therefore gated from the end of the test word to the end of the sentence in 50msecs increments without regard to word boundaries and submitted the gated sentences to independent judges who determined the gate during which the onsets of the subsequent words occurred; he reported that the judges and the experimenter concurred in 80% of cases.

Bard et al (1988) reported a large-scale study of the processing of spontaneous speech using word-level gating. They conducted two control experiments to answer possible objections to aspects of the paradigm. First, they investigated the effect on recognition outcomes of adding more left context to the gated stimulus, since it might be argued that the high rate of failures to recognise testwords might be attributable to the removal of this information source. They presented the whole of the conversation up to the onset of the test sentence used in the main experiment, as well as the gated test utterance itself, and compared the resulting responses with those made by subjects who had heard only the context provided by the gated utterance itself. While addition of prior context reduced the rate of late recognition and the average delay to recognition, it did not eradicate the phenomenon, nor alter the proportion of recognitions in categories identified in the main experiment.

In a second control experiment, Bard et al considered whether imposition of word boundary information by 'cutting' the speech signal at particular points might have been a source of bias; they therefore compared responses to segmented and unsegmented stimuli, and found only negligible differences. In the same control experiment, they showed that it was subsequent context, rather than the mere repetition of previously presented words, which contributed to the eventual recognition of stimuli.

4.2. Suitability of the Gating Paradigm for the Present Study

The foregoing discussion shows that the gating paradigm has much to offer in the context of the research questions outlined in Chapter 2. Firstly, of all the paradigms used in word recognition research, gating is the only one which can produce extensive amounts of data, both qualitative and quantitative, relating to lexical access. The extent to which the qualitative data are amenable to analysis has been demonstrated by Tyler (1984). The gating paradigm thus seems the most suitable for the investigation of the hypotheses discussed earlier. Two further points need to be considered, however: first, the treatment of words not recognised by acoustic offset, and second, the mode of response by subjects.

One aspect of the hypotheses outlined in Chapter 2 involved the nature of processing of conversational speech. Given the results of Pollack and Pickett (1963) and Bard et al.(1988) concerning the low intelligibility of some conversational tokens, the usual practice of substituting the total number of gates in the word for the isolation score might prove unacceptable because it would obscure the object of study. An alternative method needs to be adopted for the analysis of the conversational speech data. Additionally, it would be useful to have some indication of the pattern of search for words which failed to be recognised before their offsets. Fortunately, Grosjean (1985) has extended the technique in a way that makes the acquisition of such data possible, by gating not only the stimulus itself but also several subsequent words in the sentence. Therefore, when the processing of spontaneous tokens in context is examined, the post-testword gating technique will be used.

One potential disadvantage of the paradigm is the length of the experimental sessions. Although the experiment can be designed so as to avoid effects of fatigue and boredom on the part of the subjects, it seems advisable keep sessions as short as possible. Therefore, in the experiments described below, subjects will respond verbally, rather than in writing.

5. Summary

This chapter set out to describe the range of experimental paradigms available for the investigation of the word recognition process, and to justify the selection of a single technique, gating, for use in the present study. The requirements of the study were first presented: while it is necessary to be able to identify the point at which the word is recognised, the pattern of processing in the access phase is another aspect of the data which requires investigation. Although the multitude of so-called 'on-line' techniques (in which the dependent variable is generally a reaction time) provide information about the former, few can shed any light on the latter. The gating paradigm can give some indication of the time at which the word is recognised, relative to its onset, but its main advantage lies in the record of erroneous responses prior to recognition. Other researchers who have used the technique have usually assumed that the pooled responses of all subjects hearing a given word fragment represents the set of word candidates accessed by an individual hearer. In the present study, the successive presentation format, with spoken responses, will be adopted.

Chapter notes

(1) It is customary to subtract some constant amount from the recorded reaction time to account for the execution of the response: see, for example, Marslen-Wilson and Tyler (1980).

(2) Note, however, that rate is but one of the variables which distinguishes read from conversational speech: see Shockey (forthcoming).

CHAPTER FOUR

An Experiment using Read Speech

1. Introduction

In Chapters One and Two, three aspects of word recognition models were identified: the unit used for phonological access to the mental lexicon; the nature of the mechanisms designed to cope with failures to recognise words by their acoustic offsets; and the use of context in auditory word recognition. This chapter reports an experiment designed to examine all three issues. The Cohort Model (Marslen-Wilson and Tyler 1980; Marslen-Wilson 1987) is used as a theoretical starting-point, because it makes testable claims in the areas which have been identified. Three predictions of the Cohort Model, arising from the three issues just mentioned, form the basis of the work described in this thesis: the phonological access hypothesis, the early recognition hypothesis, and the context-free hypothesis. It should be noted, however, that although the research questions which are examined in this and the next chapter are framed in terms of the Cohort Model, the issues which underlie the questions are of wider relevance.

The issue of phonological access must be addressed by all models of auditory word recognition: what part(s) of the word does the hearer use to make contact with the forms in the mental lexicon, given an audible input which approximates more or less closely to the form which is assumed to be listed there? According to the Cohort Model, hearers use the initial portions of words to select a set of word candidates for further processing (1). The 1980 version of the theory (CM1) makes strong, and easily testable, predictions: the forms which are accessed in the mental lexicon will match the input exactly in terms of the phonological specification of their initial portion; and in cases where hearers enter the wrong cohort, they will never manage to recognise the word correctly. This claim is at odds with the predictions of another model, proposed by Grosjean and Gee (1987), which will be termed the Stressed Syllable model. According

to this model, it is stressed syllables, rather than word onsets, which are used to make contact with the forms in the mental lexicon:

> **Stressed syllables (and only they) are used to initiate a lexical search. (Grosjean and Gee 1987:144)**

The strong version of this theory requires that hearers should be able to distinguish stressed from unstressed syllables pre-lexically, before the word itself has been recognised. This portion of the input is then matched against stressed syllables in lexical entries, and words with suitable characteristics are accessed. If hearers are able to categorise syllables in this way pre-lexically, then the stressed syllable approach might be an efficient means of lexical access. But classification of syllables in this fashion may not be straightforward, as the acoustic features of increased amplitude and duration and fundamental frequency movement do not correlate perfectly with the location of stressed syllables. It may thus be necessary to retreat from this strong position and postulate a weak account of lexical access by stressed syllable.

A weak version of the theory is that hearers attempt to operate such a strategy but that they are frequently inaccurate in their classification of syllables in the input. When listeners classify an unstressed syllable as stressed, an unnecessary and undesirable lexical access will be initiated. When a stressed syllable is misclassified, no lexical access attempt will take place and the progress of the sentence processing operation will be impeded. Thus the efficiency of the strategy will decrease to the extent that hearers inaccurately classify an input syllable. Even within the weaker version, however, it is possible to make some predictions. It should not be the case that a subject, on hearing a word initial syllable which s/he accurately classifies in segmental terms, produces a response in which the perceived segments form part of an *un*stressed initial syllable. For example, the sequence /dai/ could be the initial portion of a number of words, such as *diver*, *dinosaur* and *diatribe* (with word-initial stress)

and *divert, direct* and *digress* (stressed on the second syllable). If a hearer perceives the sequence /dai/ and is unsure about the stress level, the weaker version of the stressed syllable theory allows a lexical access attempt to be made, but the sequence must be matched to the stressed syllable of words in the lexicon. Thus *diverse, direct* and *digress* ought not to be accessed, although items such as the initially stressed words mentioned above, as well as other words with /dai/ in their stressed syllable (e.g. *indict*) are permissible. If subjects *do* access *digress* etc, the theory could only account for it by postulating that perceived segments can only be mapped on to syllables containing full vowels. Such a modification would reduce the predictive power of the model still further; in fact, it would no longer be appropriate to call it a *stressed syllable* model (2).

Other predictions follow from the stressed syllable hypothesis. While unstressed syllables will enable subjects to access words which at best only match the input minimally in terms of phonological specification, exposure to stressed syllables will lead to the accessing of words which match the input closely. Since CM1 and the Stressed Syllable model make similarly strong predictions about the nature of the access process, they will be used as the basis of the experimental investigation of the phonological access hypothesis described in this chapter. The issue of the extent to which the mechanisms outlined in the later version of the Cohort Model (Marslen-Wilson 1987) are required to explain the experimental results will be deferred until Chapter 6.

The second issue which was discussed in earlier chapters concerned the intelligibility of words: given that various studies (e.g. Pollack and Pickett 1963; Bard and Anderson 1983; Wheeldon 1985; Bard et al.1988) have revealed that many words will remain unrecognised by their acoustic offset, how can models cope with the consequences for continuous speech recognition? Despite the evidence of the studies just cited, Marslen-Wilson (1987:80) has claimed that polysyllabic words, and indeed any words heard in constraining contexts, will be recognised before their acoustic offsets. Given the emphasis which the model places on the accurate identification of word onsets, it is crucial that support should be found for this early recognition hypothesis; for if hearers are to know where words begin, they can often do so only by determining where the preceding word ends (since word onsets in connected speech are often not explicitly marked), and this in turn can only be accomplished when that preceding word has been recognised.

The final prediction, which was termed the context-free hypothesis, concerned the use of left context in the recognition of words: according to CM1, hearers assemble word candidates on the basis of acoustic-phonetic information alone, and only subsequently take into account contextual factors; thus it is predicted that it will be possible to identify an initial phase of processing during which context effects will be absent.

The gating experiments which will be reported in this chapter examine all three hypotheses. Two experiments are described; the second is a replication of the first, using subjects with different educational backgrounds which, it was predicted, might affect the manner in which they processed words. In general, the two experiments produced similar results, although some interesting differences were observed.

To investigate the phonological access hypothesis, a lexical stress variable was introduced into the design of the experiment; trisyllabic test words bore primary stress on either the first or the second syllable. The structure of the carrier sentences was such that the test words were realised with genuine acoustic differences between stressed and unstressed syllables (for a discussion of this point, see, for example, Cutler 1986). The responses offered when these gated test words were presented to subjects were examined with respect to several dependent variables in order to assess the conflicting claims of CM1 and the Stressed Syllable model. Most of the analyses presented in relation to the phonological access hypothesis are concerned with the processing of the initial syllable, about which the two models make markedly different predictions. CM1 predicts that the responses which subjects offer after hearing the initial portion of the word should belong to the same word-initial cohort (defined in phonological terms) as the input, irrespective of whether the initial portion of the word belongs to a stressed or unstressed syllable. The Stressed Syllable model, on the other hand, appears to require that hearers will be sensitive to the stress level of syllables, or at least that, when presented with stressed initial syllables, they will be more likely to produce responses which are phonologically similar to the input than when they hear unstressed syllables; the results of this analysis, presented below, fail to support CM1 unequivocally, since there is some evidence of differential processing of the two syllable types. On the other hand, subjects show only limited ability to distinguish prelexically between stress levels. Furthermore, CM1 predicts that, after hearers have been presented with the first 150 msec of the test word, they should consistently offer responses from the same word-initial cohort as the input; while this prediction finds support in the data collected in the with-context conditions, cohort entry points for words heard in isolation are somewhat longer.

In order to examine the early recognition hypothesis, two measures of recognition were used to analyse the success rate in the identification of polysyllabic words when they were heard in full (that is, at their final gate). In general, the high rate of recognition both for words heard in context and for words heard in isolation is consistent with the model's predictions, at least for carefully articulated speech such as is used in this and in most other psycholinguistic experiments.

To test the context-free hypothesis, identical gated word-tokens were presented in three contextually-defined conditions: with neither left nor right context; with minimal left context; and with highly constraining left context. The experiments reported here do not represent the first attempt at addressing the issue of contextual preselection (see, for example, Grosjean (1980) and Tyler (1984)). Tyler's paper is of particular interest in this connection. Examining responses from a gating experiment, Tyler found that, of the data collected from the first gate of each of her with-context conditions, 5% of responses were contextually inappropriate. She used this result to support the Cohort Model's claim that contextual factors were not allowed to influence cohort initiation, arguing that, if contextual preselection were involved, *no* responses offered during these early gates would be semantically or syntactically inappropriate (1984:423). Tyler excluded her no context data from these analyses, reasoning that only data produced in the with-context conditions were relevant to the assessment of the contributions of top-down factors. In the analyses of the data from the present experiment, a somewhat different approach from Tyler's will be adopted: the responses offered in the No Context condition will be compared with those offered in the with-context conditions, on the assumption that the No Context data must reflect the likelihood of syntactically or semantically appropriate candidates occurring by chance; any significant improvement over the No Context mean in the with-context conditions must therefore be attributable to contextual factors. The results of the analyses presented in this chapter indicate that, contrary to the predictions of the Cohort Model, subjects use contextual information to influence the size and composition of the set of word candidates which is considered at the early stages of processing of words.

2. Experiment 1a

2.1. Method

2.1.1. Materials

Twenty-four pairs of trisyllabic nouns, matched for initial segment (3) and, as far as possible, for word frequency (Kucera and Francis 1967) were selected. In each pair, one word bore primary stress on the first syllable ('Syll-1 words') and the other on the second syllable ('Syll-2 words'). An example of such a pair is *carpenter* (eleven occurrences per million words of text) and *cartoonist* (four occurrences per million words of text). The test words, with word frequency marked, are listed in Appendix A.

Word frequency did not differ significantly (Syll-1 mean = 57.4, s.d. = 100.6; Syll-2 mean = 33.8, s.d. = 43.9; $t = 1.05$, $p = 0.2981$). The frequency match is probably better than it might seem on the basis of these group means: the largest discrepancy was between *company* (453 occurrences per million) and *computer* (18 occurrences per million). It seems likely that the word *computer* is now more common than its 1967 frequency score would suggest. When this pair was removed from the calculation, the difference between the two groups was even less significant: $t = 0.38$, $p = 0.7057$, with means of 40.22 and 34.52, and standard deviations of 56.27 and 44.72, for Syll-1 and Syll-2 words, respectively.

For each test word, cohort size (that is, the number of words in the lexicon sharing the first two phonemes of the test word) was computed automatically using a computer-readable version of the Oxford Advanced Learners' Dictionary (which contains 18,000 head entries, each with a broad phonetic transcription). Although this variable had not been explicitly controlled, cohort sizes for words stressed on either the first or the second syllable were extremely similar: for words stressed on the first syllable, mean = 200, s.d. = 157; for words stressed on the second syllable, mean = 204, s.d. = 174 ($t = -0.08$, $p < 0.938$).

Contexts for the test words were constructed according to principles similar to those adopted by Grosjean (1980). For the example used above, the test sentences were as follows:

> (4.1) Because he liked making things, Tom became a carpenter after leaving school
>
> (4.2) Because she liked drawing things, Ann became a cartoonist after leaving school

Each test word was embedded in the second clause of a two-clause sentence whose structure was such that the second clause could stand alone as a well-formed sentence. Thus, a single sentence provided two context conditions: the Long Context condition, which comprised the whole sentence, and the Short Context condition, derived from the Long Context sentence by deleting the first clause. For each pair, every effort was made to match features of grammatical structure and prosody, subject to the constraint that the sentences should offer similar contextual cues to the identity of the test word (see below). The test word was preceded by between four and six syllables in the Short Context sentence and was followed by a prepositional or adverbial phrase, so that the test word should not be subject to any sentence final effects such as lengthening. The sentences remained semantically and syntactically well-formed even after the removal of the final phrase. The test sentences are presented in Appendix B.

The sentences were then pretested, using a visual cloze procedure, to determine the extent to which they predicted the test word: the Long Context and Short Context sentences, up to and including the word before the test word, were presented to two groups of ten subjects (4) who were asked to complete each sentence with a single word. Their responses were then compared with the intended word by an independent panel of three judges, and scored according to the metric suggested by Marslen-Wilson and Welsh (1978): 1 for a word identical to the target, 2 for a synonym, 3 for a related word and 4 for a totally unrelated word. From these scores, the mean for each of the 96 contextual stimuli was computed. The aim was to produce sentences whose mean for the highly constraining Long Context condition was less than two and for the less constraining Short Context condition was between three and four. The sentences were adjusted and re-tested (using a new set of judges) until these criteria were met.

The final versions of the Long Context sentences, including the sentence-final prepositional or adverbial phrase, were recorded by a male speaker of Scots English. These stimuli were then digitised using a PDP 11/40 computer and were manipulated using the ILS signal processing package (Signal Technology Inc. 1985). Visual and auditory information were used to locate four points in each stimulus: the onset of the first clause, the onset of the second clause, and the onset and offset of the test word (5).

Thus, from a single recorded sentence, three stimuli were produced:

Long Context - measured from the onset of the first clause to the offset of the test word. Example: *Because she liked drawing things, Ann became a cartoonist.*

Short Context - measured from the onset of the second clause to the offset of the test word. Example: *Ann became a cartoonist.*

No Context - measured from the onset of the test word to its offset. Example: *cartoonist.*

The lengths of the test words differed across the two stress categories: Syll-1 mean 515.92 msec, S.D. 92.46; Syll-2 mean 579.71 msec, S.D. 103.76; $t(1,46) = 2.25$, $p = 0.0292$. Such a difference has been noted elsewhere in the literature (Cutler and Clifton 1984).

The stimuli were gated automatically by 50 msec increments and generated at every experimental session by the PDP11/40 computer (6). This approach enabled the efficient collection of subjects' confidence ratings (see Procedure). In addition, subjects were thereby able to control, albeit indirectly, the speed with which they heard subsequent gates and therefore the amount of time available for response, since the presentation of the next stimulus was initiated only after the subject had pressed the key to indicate confidence rating.

2.1.2. Design

The design was as follows: Stress (2 levels: Syll-1, Syll-2) X Context (3 levels: No Context, Short Context, Long Context). Although each subject heard every test word, no individual heard a given word in more than one Context condition. Subjects were nested in groups defined by the constraint just described. Words were nested in Stress but crossed with Context; in all other cases, a crossing relationship existed between variables.

2.1.3. Subjects

Eighteen undergraduates, native speakers of Scots English attending a wide variety of courses at the University of Edinburgh, volunteered to take part in the experiment. The mean age of the subjects was 19.2 years. Each subject participated in a single experimental session lasting approximately 75 minutes, with a break about half-way through the session.

2.1.4. Procedure

Subjects were seated before the computer screen in a sound-proofed room; stimuli were presented through earphones. Subjects were told to respond aloud after each presentation of a stimulus with their response as to the identity of the gated word, which was the final word in the sentence. These responses were recorded and the resulting tape was used to check the written record of responses made by the experimenter at the time of the experiment. In addition, subjects were to press a key to indicate their confidence in their response, on a 1 - 5 scale (1 for very unsure, 5 for absolutely certain). Their key-press triggered the presentation of the next test item. Three practice items (presented in Appendix C) preceded the experimental stimuli, which were grouped in blocks according to the Context variable; the order of presentation of blocks was counterbalanced. Subjects took a break after either the first or the second block, according to their own preference. An appropriate visual display appeared on a computer terminal one second before the presentation of a fresh stimulus and half a second before the onset of each successive gate.

2.2. Results

Following Clark (1973), the data examined in most of the subsequent analyses were input to by-subjects and by-materials ANOVAs, to yield values for F1 and F2 respectively. These values were then used to compute MinF′, a statistic which enables the researcher to make generalisations about the applicability of a result both to further subjects and to further materials.

Unless otherwise stated, interactions failed to reach significance in the analyses described in the following sections.

2.2.1. The Phonological Access Hypothesis

The data examined in relation to the phonological access hypothesis were drawn almost exclusively from the first few gates which subjects heard. Since the claims of CM1 relate to the processing of the initial portion of the word up to and including the first vowel, the gate containing this segment was identified in the following manner. The experimenter selected a point mid-way through the second syllable and two phoneticians listened to all the gates corresponding to the material up to and including that point. They provided a narrow phonetic transcription of all the gates they heard and their transcriptions were used to identify the gate containing the initial vowel. The initial vowel occurred at a later gate for words stressed on the first syllable than for words stressed on the second syllable (mean for Syll-1 words, 3.5833, s.d. 0.9286; mean for Syll-2 words, 3.00, s.d. 0.8341; $t = 3.44$, d.f. $= 23$, $p = .0022$).

When the gate containing the end of the initial vowel had been identified in this manner, several analyses were performed. Firstly, the responses at this gate were examined to see whether the prediction of the strong version of the stressed syllable model, that hearers are sensitive to the stress level of syllables, would be borne out. There is some limited evidence for such a skill.

In a second analysis, the responses offered at the gate containing the end of the first vowel were compared with the input words to see whether they were drawn from the same word initial cohort as the input. The finding that stressed syllables are more efficient than unstressed syllables at constraining the responses of subjects is difficult to explain in terms of the 1980 version of the Cohort Model. A further respect in which the predictions of the model are not borne out is in subjects' recovery from initial failures to enter the cohort at an early stage of processing. Even though such failures are represented in the data, in most cases, cohorts were entered and words were successfully identified by acoustic offset.

Finally, the cohort model's prediction that, after 150 msec of the stimulus have been heard, subjects will be offering responses drawn from the correct cohort, is tested; while this claim finds support in the with-context conditions, it fails to do so for words heard in isolation.

2.2.1.1. Sensitivity of subjects to the stress pattern of input

A prerequisite of the strong version of the stressed syllable model is that hearers should be able to distinguish between stressed and unstressed syllables at a pre-lexical stage of processing. If they fail to do so, they will be unable to identify stressed syllables and to use these to gain access to the mental lexicon. If no such ability is found, and the strong version of this stressed syllable model must be rejected, it may still be possible to find evidence for a weaker version of the account, as was discussed above: if hearers are poor at distinguishing stressed from unstressed syllables pre-lexically, they may initiate lexical access attempts too frequently, that is, on the basis of syllables which are unstressed as well as syllables which are stressed, but in either case they will access lexical entries which conform to the input as far as the stressed syllable is concerned. Such a possibility will be considered in the next section, but the analyses about to be presented are concerned with examining the strong version of the theory.

Responses at the gate containing the end of the first vowel were examined to determine the percentage of words whose initial syllable was stressed. (Monosyllabic words were considered to be initially stressed.) For the claims of the strong version of the stressed syllable model to find support, *all* responses offered during these gates in Syll-1 words should have stressed initial syllables, while *no* responses offered during these gates of Syll-2 words should have stressed initial syllables.

Table 4.1: Percentage of responses with initial stress which were offered at the gate containing the end of the first vowel (Experiment 1a)

CONTEXT	none	short	long	mean
STRESS				
Syll-1	86	95	97	(93)
Syll-2	70	33	11	(38)
Mean	78	64	54	

Table 4.2: ANOVA for Table 4.1, related to Stress and Context

Source	F1	F2	MinF′
Stress	(1,17) 395.89 **	(1,46) 265.75 **	(1,58) 159.01 **
Context	(2,34) 18.89 **	(2,92) 19.79 **	(2,97) 9.66 **
Stress X Context	(2,34) 83.00 **	(2,92) 43.72 **	(2,122) 28.64 **
** $p < .01$			

Both main effects were highly significant. In the case of Context, the conditions for the two with-context means did not differ reliably, but the No Context mean was significantly higher than the other two ($p < .01$, by-subjects Scheffé test). Stress and Context interacted, such that although the means for Syll-1 words were similar, the Syll-2/Syll-1 difference, and the difference between the means in the Syll-2 condition, were significant ($p < .01$, by-subjects Scheffé test).

At first sight, it might appear that subjects in the Syll-1/with-context conditions were exhibiting a strong awareness of the stress level of the initial syllables of the test word. Closer inspection of the responses in these conditions, however, reveals that the high number of responses with word-initial stress was largely accounted for by the high rate of recognition of test words during early gates in these conditions (7): in the Syll-1 condition, such responses will clearly increase the number of lexical items stressed on the first syllable. In the present analysis, the number of responses in the Syll-1/with-context conditions which were not the test word itself but which were stressed on the first syllable is very small indeed. This interpretation is confirmed by inspection of the corresponding responses in the Syll-2 condition: for example, of the 89% of responses with unstressed initial syllables, almost all were the test word itself.

The presence of context makes the interpretation of these data with respect to the strong syllable hypothesis difficult.

The No Context responses provide evidence which is uncontaminated by the effects of higher-level processing. In both conditions, the percentage of responses which were initially stressed is high, contrary to the predictions of the stressed syllable model, which would require that the proportion should be low in the Syll-2 condition. This high percentage of initially stressed words is accounted for in part by the tendency of subjects to respond at the early gates with monosyllabic words (this trend is apparent in the examples of response sets provided in Grosjean's 1980 paper). However, if the analysis is confined to the polysyllabic responses in the No Context condition, the results are no better than those from the full data set: in the Syll-1 condition there were 101 such items, and 84 of these were initially stressed; while in the Syll-2 condition, 44 of the 88 polysyllabic responses were initially stressed.

Overall, the mean percentage of correct assessment of stress level (i.e. initially stressed responses in the Syll-1 condition, and initially unstressed responses in the Syll-2 condition) is only 58%; this level of accuracy does not appear to be sufficient to support the claims of the strong version of the stressed syllable model. However, it is still possible to sustain the stressed syllable model in a weak form even with this poor ability on the part of subjects to distinguish stressed and unstressed syllables, provided that the system is able to tolerate a large number of false alarms: in particular, in the Syll-2 condition, a large number of lexical access attempts are initiated which would prove fruitless. To find evidence which confirms or denies the existence of the stressed syllable strategy it is necessary to conduct further analyses.

2.2.1.2. Phonological match between initial portion of input and response

This analysis was confined to responses offered at the gate containing the end of the first vowel. CM1 predicts that, by this stage, hearers should be offering responses which match closely with the input for this initial portion at least, irrespective of whether the signal they hear corresponds to a stressed or an unstressed syllable. The Stressed Syllable model, by contrast, predicts that, while a relatively close match will be found for words with stressed initial syllables, this will not be the case when initial syllables are unstressed.

Table 4.3: Percentage of responses at the gate containing the end of the first vowel (Experiment 1a) which match the input for that initial portion

CONTEXT	none	short	long	mean
STRESS				
Syll-1	61	80	94	(78)
Syll-2	48	67	85	(67)
Mean	55	74	90	

Table 4.4: ANOVA for Table 4.3, related to Stress and Context

Source	F1	F2	MinF′
Stress	(1,17) 28.40 **	(1,46) 4.53 *	n.s.
Context	(2,34) 30.98 **	(2,92) 23.72 **	(2,111) 13.43 **
** $p < .01$	* $p < .05$		

In this analysis, words stressed on the first syllable elicited more responses which agreed with the input in initial portion than did words stressed on the second syllable; while this result just failed to reach significance in the MinF′ analysis, the pattern of results is not entirely consistent with the predictions of CM1, which does not predict a different pattern of responses to words in the two Stress categories.

At first sight it appears that the results of this analysis offer more support to the stressed syllable hypothesis. However, closer examination of the responses reveals some interesting evidence which is relevant to this issue. It was noted above that even the weak version of the stressed syllable hypothesis made predictions about the kinds of responses which subjects might produce. In particular, when subjects are presented with the initial syllable of a word, they should not map the segments which they perceive on to an unstressed syllable in the lexicon (for example, the /tɛ/ of *temperature* should not elicit *temptation*). Yet there are several instances of just such behaviour in the data for the no-context condition. It is the Syll-1 responses which are of particular interest in connection with this argument. Of the 101 polysyllabic responses offered in the Syll-1/no context condition, 17 had unstressed initial syllables and 7 of these cases involved the mapping of the segments which were presented onto unstressed syllables. Thus, the data from this experiment, while indicating that stressed syllables are more intelligible than unstressed, does not entirely support the account of access proposed by writers such as Grosjean and Gee.

The differences among the three Context means was highly significant ($p < .01$, by-subjects Scheffé test). This result is not predicted by CM1 since processing during the initial portions of words is supposedly confined to lexical access on the basis of acoustic-phonetic information; the importance of this fact will be considered in Section 2.2.3.

According to the predictions of CM1, the 239 instances in which subjects failed to enter the appropriate cohorts should never result in successful identifications of the words in question. However, examination of the continuing responses of subjects to these word tokens reveals that this is not the case. Table 4.5. summarises the continuing processing of these words. For each cell in the design, four figures are presented: the number of words whose cohorts were not entered by the end of the word's initial C(C)V; the number of these words which were eventually identified correctly; the number of the remaining words which, though they were never recognised correctly, elicited responses at the final gate which were drawn from the same cohort as the target; and the number of words which, by the final gate, elicited responses from a different cohort.

In the majority of cases, hearers subsequently entered the correct cohort and indeed frequently identified the word accurately. There are only four instances (out of a possible 288 in the full experiment) where subjects were unable to enter the correct cohort by the offset of the word. In two cases (both from the No Context condition), subjects identified the initial consonant correctly but not the vowel. The remaining instances seemed to arise through the subjects' misanalysis of word boundaries. For example, in one case, the subject, hearing the stimulus

(4.3) They had a *parliament*

failed to detect a word boundary after *a*, and after producing the response *apartment* several times, suggested *appalling* at the final gate.

Table 4.5: subsequent processing of those words whose cohorts were not entered by the end of the first C(C)V

CONTEXT		none	short	long
STRESS				
	initial failures to enter cohort	55	28	9
Syll-1	subsequent successful identifications	51	26	9
	CE, but failure to identify	2	1	0
	failure to enter cohort by acoustic offset	2	1	0
	initial failures to enter cohort	75	48	22
Syll-2	subsequent successful identifications	73	46	21
	CE, but failure to identify	2	2	0
	failure to enter cohort by acoustic offset	0	0	1

In summary, the results of this analysis demonstrate that failure to enter the appropriate word-initial cohort in the early stages of the processing of the word in no way prevents hearers from later identifying the word. The set of candidates which were initially accessed sometimes diverged from the idealised representation, but hearers were subsequently able to recover from errors and narrow in on the intended word.

2.2.1.3. Cohort entry points

The analyses presented in Section 2.2.1.2. showed that, in a substantial proportion of cases, subjects were failing to enter the appropriate word-initial cohort by the time that they were hearing the word's initial C(C)V. Perhaps this is simply due to some time-lag in hearers' ability to utilise acoustic-phonetic information. The analysis presented in this section seeks to determine how much of the word subjects needed to hear before they were able to enter the word-initial cohort. For CM1, this duration is estimated at approximately 150 msec (Tyler 1984).

The gate at which subjects entered the word-initial cohort (that is, began to offer responses agreeing with the testword in the initial C(C)V), and did not subsequently offer as responses words which were not members of that cohort, was calculated; this measure will henceforth be termed the Cohort Entry (CE) point.

Table 4.6: Mean number of 50msec gates heard before Cohort Entry (Experiment 1a)

CONTEXT	none	short	long	mean
STRESS				
Syll-1	4.3	2.9	2.0	(3.1)
Syll-2	4.4	2.9	2.0	(3.1)
Mean	4.4	2.9	2.0	

Table 4.7: ANOVA for data in Table 4.6. related to Stress and Context

Source	F1	F2	MinF′
Stress	n.s.	n.s.	---
Context	(2,34) 18.56**	(2,92) 46.70**	(2,122) 37.43**
** p < .01	* p < .05		

Note that although in the with-context conditions the predictions of CM1 are borne out, in the No Context condition the Cohort Entry point was somewhat later than the 150 msec (three gates) which is the latest estimate of the duration of input required to assemble the set of word candidates.

Although a stress difference was found when the dependent variable was the number of responses matching the C(C)V stimulus, this difference is not apparent in the present analysis: the stress means are identical. That is, it appears that any temporary difficulty subjects were experiencing when processing unstressed initial syllables has been washed out by this point. In the with-context conditions this is largely due to the presence of contextual cues. In the No Context condition, the absence of the effect must be due solely to the subjects' use of continuing acoustic-phonetic information. It is likely that the acoustic event which enabled cohort entry when subjects heard the fourth to fifth gate of Syll-2 stimuli was the second (i.e. stressed) syllable of these items. That is, because the first syllable of the Syll-2 items would have been shorter than that of the Syll-1 items, four or five gates of these items was more likely to include part of the second syllable. Thus, it is possible that a stress effect similar to that revealed in the previous analysis is obscured here, because of the fact that the dependent variable examined is not sensitive to the differing lengths on the initial syllables of the test word.

The Context effect was highly significant. This was due not only to the later-than-predicted CE points for words heard in isolation, but also to the very early CE points in the Long Context condition. A post hoc comparison of means by Scheffé test showed that all three Context means differed significantly ($p<.01$, by subjects). Note that co-articulation effects, which might be claimed to assist subjects in the with-context conditions, cannot be solely responsible for the effect, since the difference between the means in the Short and Long Context conditions (which contained identical co-articulatory cues) is as marked as that between the No Context and Short Context conditions. In any case, subsequent analyses will show that the syntactic and semantic characteristics of the set of word candidates display effects of context; indeed, it is this ability which in part accounts for the very early CE points in the Long Context condition, since in this condition it was not uncommon for words to be recognised at the first gate. The context result will be discussed in more detail below.

2.1.1.4. Summary of evidence relating to the Phonological Access Hypothesis

The analyses presented in this section addressed the issue of the use of acoustic-phonetic information to gain access to the representations in the mental lexicon. Two conflicting claims were assessed: CM1 predicted that the initial syllables of words were the basis of phonological access, while the Stressed Syllable model assigned this role to stressed syllables.

The predictions of CM1, that hearers would enter the word-initial cohort after hearing some 150 msec of the testword and that there would be a close match, at a phonological level, between the initial portion of input and target, were not supported: the results were in general more consistent with some form of stressed syllable model. Subjects displayed a limited ability to distinguish stressed from unstressed syllables before recognising the words in which they occurred. More importantly, there was evidence that stressed syllables were more efficient than unstressed at eliciting phonologically appropriate responses (though it was argued that the occurrence of responses in which the stimulus syllable was mapped onto an *un*stressed syllable in the response was inconsistent with the stressed syllable model's claims). However, when the dependent variable was the point at which subjects actually began to offer appropriate responses consistently, the advantage associated with words stressed on the first syllable disappeared.

2.1.2. The Early Recognition Hypothesis

The Early Recognition Hypothesis which was discussed in the introduction stated that, in the majority of cases, polysyllabic words, and all words heard in context, would be recognised before their acoustic offsets. In earlier gating studies (e.g. Grosjean 1980; Tyler and Wessels 1983; Salasoo and Pisoni 1984) a number of dependent variables have been used to monitor the process of word recognition. Most studies have noted the point at which subjects begin to offer the target word as their response, and do not subsequently change their response: Grosjean (1980) termed this the *isolation point*. When subjects provide a rating of their level of confidence in their response, a further dependent variable becomes available: the point at which a subject reaches (and subsequently maintains) a predefined level of confidence was named the *recognition point* by Grosjean. In this section, both types of data will be considered, and Grosjean's terminology will be adopted.

2.1.2.1. Percentage of words successfully isolated by the final gate

The first dependent variable which was examined was the percentage of words in each cell isolated (see definition above) by the final gate. Cell means are presented in Table 4.8. and the corresponding ANOVA results in Table 4.9.

The level of isolation was extremely high in all cells. As the Cohort Model predicts, the majority of words are isolated by the final gate. This is true not only in the case of words heard in context, but also of words heard in isolation. No Stress difference was found, and the weak Context effect was significant in the by-subjects analysis only.

Table 4.8.: Mean percentage of successful isolations by the final gate (Experiment 1a).

CONTEXT	none	short	long	mean
STRESS				
Syll-1	96	97	100	(98)
Syll-2	98	99	99	(99)
Mean	97	98	100	

Table 4.9: ANOVA for Table 4.8, related to Stress and Context

Source	F1	F2	MinF′
Stress	n.s.	n.s.	---
Context	(2,34) 4.50 *	n.s.	---
** $p < .01$	* $p < .05$		

2.1.2.1.1. Percentage of words successfully recognised by final gate

The term *recognition*, for the purposes of the analysis presented here, is used in the sense first suggested by Grosjean (1980). That is, a subject is deemed to have recognised a word when, in addition to having isolated the word, s/he maintains a confidence rating at or above a criterial level (here, the level selected was 4 on the 1 - 5 scale (8)). The mean percentages of successful recognitions are displayed in Table 4.10, with the corresponding F-values in Table 4.11.

The Stress means did not differ significantly. This is not surprising in view of the finding, discussed above, that stress differences were not detectable at the cohort entry point. The differences among all three Context conditions was, however, significant ($p < .01$, by-subjects Scheffé test): although the level of recognitions was high in the with-context conditions, 12% of words heard in isolation failed to be recognised with high confidence by the final gate.

Table 4.10: Mean percentage of successful recognitions at the final gate (Experiment 1a)

CONTEXT	none	short	long	mean
STRESS				
Syll-1	85	94	100	(93)
Syll-2	91	98	99	(96)
Mean	88	96	100	

Table 4.11: ANOVA for Table 4.10, related to Stress and Context

Source	F1	F2	MinF´
Stress	n.s.	n.s.	---
Context	(2,34) 14.51 **	(2,92) 9.23 **	(2,118) 5.64 **
** $p < .01$	* $p < .05$		

2.1.2.2. Summary of evidence relating to the Early Recognition Hypothesis

In this section, two measures of retrieval have been considered: the isolation point and the recognition point. Even according to the latter, more conservative measure, the majority of words heard in context are recognised before their acoustic offsets, supporting the prediction of the Cohort Model. The high level of isolations and recognitions in the No Context condition suggests that the speech sample used is highly intelligible. The manipulation of the stress variable has apparently had no effect: if the unstressed initial syllables of Syll-2 words cause any processing problems for hearers, this effect is no longer discernible by the end of the word.

2.1.3. The Context-Free Hypothesis

According to the Context-Free hypothesis, no effects of syntactic or semantic context should be discernible in the responses offered in the gating task while subjects are hearing the first three gates (50, 100 and 150 msec) of the stimulus, since during this lexical access phase subjects do no more than assemble the set of word candidates on the basis of the acoustic-phonetic properties of the input.

The dependent variables which will be examined in relation to this question are as follows:

(i) the grammatical category of responses: the model predicts that in the three context conditions, the same proportions of grammatical categories will be represented in responses;

(ii) the semantic appropriateness of responses: the model predicts that, when the responses collected in the No Context condition are compared with those collected in a with-context condition, the responses collected will be equally likely to be semantically appropriate to the sentence introduction heard in the with-context condition;

(iii) the type-token ratio of responses: the model predicts no difference in the number of distinct word types offered as responses in each condition.

2.1.3.1. Grammatical category of responses

It was pointed out above (Section 2.1.4.) that subjects were informed that the gated word was the last word in the sentences heard in the with-context conditions. Since the gated word was immediately preceded by a determiner (*a* or *the*), its only possible word class was 'noun' (9). Therefore, in the present analysis, the dependent variable was the percentage of valid responses which numbered 'noun' among their possible word classes. Results are presented for data collected at gates 1, 2 and 3 (50, 100 and 150 msec). (See Appendix D for analyses relating to responses collected at gate 1 only; for all practical purposes, results were identical to those presented in the text below.)

There was some tendency, apparent in the F1 analysis, for subjects to respond with nouns more often when the target was stressed on the first syllable than when it was stressed on the second syllable. This may reflect subjects' intuitive awareness of the tendency for nouns to be stressed on the first syllable and verbs to be stressed on the second syllable (Fudge, 1984).

Table 4.12: Mean percentage of responses collected at gates 1, 2 and 3 with appropriate grammatical category (Experiment 1a).

CONTEXT	none	short	long	mean
STRESS				
Syll-1	74	90	98	(87)
Syll-2	71	84	94	(83)
Mean	73	87	96	

Table 4.13: ANOVA for data in Table 4.12 related to Stress and Context.

Source	F1	F2	MinF´
Stress	(1,17) 5.66**	n.s.	---
Context	(2,34) 41.10 **	(2,92) 18.89 **	(2,124) 12.94 **
** p < .01	* p < .05		

The Context effect was significant, contradicting the Cohort Model's claim that, within the lexical access phase, no effects of syntactic context should be observed. In a by-subjects Scheffé test, it was found that all three context means differed at the $p < .01$ level of significance.

2.1.3.2. Semantic appropriateness of responses

The following pair of analyses compare the two with-context conditions in turn with the No Context condition. As discussed above, the underlying assumption of these analyses is that the responses offered in the No Context condition reflect the possibility of a given response occurring by chance, taking into account the constraining influence of a particular phonetic string which characterises the input. If the with-context conditions produce responses which are reliably more appropriate to context than this quasi-randomly generated set, then this difference must be attributable to the effects of top-down information alone. Since the data for these analyses were culled from responses made during the lexical access phase (first three gates) the Cohort Model would predict that no such difference should be found.

Semantic appropriateness of responses was assessed in the following manner. For each of the two comparisons (No Context versus Short Context and No Context versus Long Context) a separate pair of judges was asked to evaluate the appropriateness of each response made in the early gates of each condition in relation to the sentence introduction heard in the with-context condition. That is, for the first analysis (No Context versus Short Context), No Context responses and Short Context responses were evaluated in relation to the Short Context sentence, while in the second analysis (No Context versus Long Context) No Context and Long Context responses were evaluated in relation to the Long Context sentence. The resulting scores then formed the basis of the two sets of ANOVAs presented below.

2.1.3.2.1. Suitability of responses made in No Context and Short Context conditions as completions of Short Context sentences

No Context and Short Context responses were compared in the manner described in the previous section. Cell means and results are shown in Tables 4.14 and 4.15 respectively.

Table 4.14: Percentage of responses offered at gates 1, 2 and 3 in the No Context and Short Context conditions which were semantically appropriate completions of the Short Context sentences (Experiment 1a).

CONTEXT	none	short	mean
STRESS			
Syll-1	31	83	(57)
Syll-2	24	79	(52)
Mean	28	81	

Table 4.15: ANOVA for Table 4.14 related to Stress and Context

Source	F1	F2	MinF´
Stress	n.s.	n.s.	---
Context	(1,17) 401.2 **	(1,46) 126.99 **	(1,63) 95.87 **
** p < .01	* p < .05		

The Stress means did not differ significantly. As far as the Context effect was concerned, the No Context condition produced fewer appropriate responses than the Short Context condition. This result demonstrates that contextual factors were implicated in the selection of the set of word candidates.

2.1.3.2.2. Suitability of responses made in No Context and Long Context conditions as completions of Long Context sentences

In this section, the results of the comparison of the No Context and Long Context data are presented.

Table 4.16: Percentage of responses offered at gates 1, 2 and 3 in the No Context and Long Context conditions which were semantically appropriate completions of the Long Context sentences (Experiment 1a).

CONTEXT	none	long	mean
STRESS			
Syll-1	5	82	(44)
Syll-2	7	86	(47)
Mean	6	84	

Table 4.17: ANOVA for Table 4.16, related to Stress and Context

Source	F1	F2	MinF′
Stress	n.s.	n.s.	---
Context	(1,17) 1177.09 **	(1,46) 595.43 **	(1,61) 395.41 **
** p < .01	* p < .05		

Although the Stress effect failed to reach significance in any analysis, there was once again evidence of the effect of subjects' use of context in the selection of the set of word candidates considered. That is, there was a greater likelihood of words appropriate to the Long Context sentence introduction being offered in the Long Context condition than in the No Context condition. These results provide further support for the hypothesis that hearers use context actively to constrain the set of word candidates which are considered during lexical access.

2.1.3.3. Type-token ratio of responses

The Cohort Model claims that, regardless of the context in which a word is heard, only bottom-up information is available to the word recognition mechanism during the lexical access phase. Thus, under the assumption discussed earlier, whereby the responses of subjects, when pooled, should reflect the structure of the set of word candidates assembled by a single subject, the data from the three Context conditions should produce roughly equivalent type-token ratio means.

For the analysis described in this section, responses from each subject hearing a given word in a given context condition were pooled. Type-token ratios for each word in each condition were then computed (number of word types divided by number of word tokens). These ratios were input to a single, two-way ANOVA. Because of the nature of the analysis, only the by-materials ANOVA was run.

Table 4.18: Type-token ratio of responses in first three gates (Experiment 1a)

CONTEXT	none	short	long	mean
STRESS				
Syll-1	.73	.47	.26	(.49)
Syll-2	.65	.46	.25	(.45)
Mean	.69	.47	.26	

Table 4.19: ANOVA for Table 4.18, related to Stress and Context

Source	F
Stress	n.s.
Context	(2,92) 87.87 **
** p < .01	* p < .05

As in the other analyses, the Stress means did not differ reliably. The significant Context effect failed to support the prediction of the Cohort Model. That is, the size of the set of word candidates offered as responses was influenced by the presence or absence of context: where more contextual information was available, fewer distinct words candidates were offered. According to Scheffé tests, all three means differed at the 1 percent level of significance. The No Context condition produced a higher type-token ratio than Short Context, whose mean was in turn higher than that for Long Context. In terms of the individual hearer, then, this result suggests that fewer word candidates are activated in the presence of contextual information.

2.1.3.4. Summary of evidence relating to the Context-Free Hypothesis

The prediction of the Cohort Model with respect to the role of contextual factors in the processing of the initial 150 msec of a word is clear: no effects of context, syntactic or semantic, should be detected. However, the results of the analyses just presented are inconsistent with this prediction: the set of responses offered by subjects in the with-context conditions are more syntactically and semantically constrained, and consist of fewer distinct word types, than the set of responses offered when no preceding context is available to the hearer.

2.1.4. Summary and Discussion of Experiment 1a

The results of Experiment 1a are summarised in Table 4.20. Experiment 1a set out to test three predictions arising from the descriptions of CM1. The first hypothesis which the experiment was designed to test concerned the manner in which acoustic-phonetic information is utilised; specifically, which part of the word is used to trigger lexical access. Two candidates were considered: CM1 assigns this role to the initial portions of words, while an alternative view is that stressed syllables are used to generate word hypotheses. Experiment 1a provides somewhat more support for stressed syllables in this role: subjects were able to process words more efficiently at the lexical access stage if they were stressed on the first syllable.

Table 4.20: Summary of the results of Experiment 1a

NOTE:
- the symbol * indicates effects significant for F1, F2 and MinF´
- analyses 1 to 3, phonological access hypothesis; analyses 4 and 5, early recognition hypothesis; analyses 6 to 9, context-free hypothesis.

Dependent Variable	Stress	Context	Interaction
1. Percentage of responses with appropriate stress level in initial syllable	*	*	*
2. Percentage of responses which matched test word in initial C(C)V	F1 & F2	*	
3. Cohort Entry Points		*	
4. Percentage of successful isolations by final gate		F1 only	
5. Percentage of successful recognitions by final gate		*	
6. Grammatical category of responses 1, 2 & 3	F1 only	*	
7. Semantic appropriateness of no-context vs short-context responses		*	
8. Semantic appropriateness of no-context vs long-context responses		*	
9. Type/token ratio of responses		*	

The second hypothesis concerned the recognition of polysyllabic words and of words heard in context. For the efficient operation of the Cohort Model, which relies heavily on the accurate perception of word onsets, it is important that words heard in context should be recognised by their offset, so that the onset of the subsequent word can be located and processed. Such a mechanism is required because of the well-attested paucity of word boundary cues in connected speech. In the sample of speech used, the majority of words heard in context were indeed recognised by the final gate, according to both measures of recognition adopted. It might appear that the no context results fail to support the early recognition hypothesis: although the less conservative measure of recognition indicated that almost all of these tokens were recognised by the final gate, when the dependent variable was sensitive to the confidence rating provided by subjects, 12% of words remained unrecognised by the final gate. However, since the test words were originally produced in context, it is perhaps not surprising that when they were excised from context recognition was less than perfect. In general, the results of Experiment 1a support the early recognition hypothesis.

However, no support was found for the context-free hypothesis. The cohort model claims that the process of lexical access is driven purely by acoustic-phonetic considerations, irrespective of the presence or absence of contextual information. The evidence of Experiment 1a suggests that this claim is false: the structure and size of the set of responses offered during what CM1 claims is the access phase varied with the different context conditions in which the identical word tokens were available. Three types of evidence were presented. First, when words were heard in context, there was a significantly greater likelihood, during the early gates, of subjects responding with a syntactically and/or semantically appropriate word than if the word were heard in isolation. Second, subjects offered fewer word candidate types during the lexical access phase when words were heard in context than when they were heard in isolation, suggesting that high-level syntactic and semantic factors were constraining the size of the candidate set. Third, analyses relating to the phonological access hypothesis showed that contextual factors were implicated in what should have been, according to the predictions of the CM1, a process relying on bottom-up information alone. Contextual effects were found in data sampled from the first three gates (150 msec) and when the analysis was confined to the first gate (50 msec). This result points either to a form of strong interaction between the lexical and syntactic/semantic processors, or to a weak interaction operating at intervals of less than 50 msec. The data from Experiment 1 cannot distinguish between these alternatives.

Of the three issues which Experiment 1a set out to examine, the one which has been least satisfactorily resolved is the phonological access question. The literature on phonological reduction which was examined in Chapter 2 seemed to suggest that unstressed syllables would pose serious problems for hearers, yet the evidence that they do so is somewhat unconvincing. Although the stress effect was detected in some of the analyses, a greater decrement in performance associated with unstressed syllables was anticipated based on the evidence of the literature reviewed in Chapter 2. Hearers had evidently recovered from any problems associated with the absence of stress well before the end of the word, since there was no stress effect found for the isolation point or recognition point analyses. (The absence of a stress effect in the cohort entry point analysis was attributable, it was suggested, to an artefact of the gating procedure.) If unstressed syllables constitute input which is less useful (in terms of phonological factors such as vowel quality) or of poorer quality (in terms of phonetic factors such as duration and amplitude) than stressed syllables, as the studies in Chapter 2 suggest, why should the subjects in Experiment 1a process them with such apparent ease? One possibility is that the unstressed syllables of clearly articulated, read speech do not, after all, introduce a sufficient decrement in acoustic-phonetic quality to result in the sort of ambiguity which is required to test the hypothesis. Further consideration of this explanation will be deferred until Chapter 5; first, other possibilities will be explored.

A second explanation of the smaller-than expected effects of manipulating stress pattern might be that the effects of context, described above, were so powerful that the relative indeterminacy of the acoustic signal did not present significant barriers to lexical access. That is, if the notion of bottom-up autonomy of word candidate selection is abandoned, as the results of Experiment 1a suggest it should be, then word hypotheses may be produced on the basis of contextual (as well as acoustic-phonetic) factors; the relatively strong constraints of the top-down factors might have produced such a small set of word candidates that even ambiguous bottom-up information could quickly reveal the target word. This, however, can only be part of the explanation, since the absence of stress effects also extended to the no context condition in the relevant analyses.

A third possibility, which will be considered in Experiment 1b, is that the subjects who took part in Experiment 1a were exceptionally skillful bottom-up processors. The subjects who took part in the experiment were - in common with most individuals participating in word recognition experiments - undergraduates; we might therefore assume that they spent relatively large amounts of time engaged in reading. It has frequently been remarked that the high level of literacy of undergraduates, who are the most frequently used source of experimental subjects, may lead to biassed results; see, for example, Brown and Yule (1983) and Linell (1982). There are several ways in which undergraduates' reading expertise might affect their processing of spoken language. For example, in visual word recognition experiments, it has been shown that the beginnings of words are more informative than their middles or ends, and that readers therefore focus attention on this part of the word (Bruner and O'Dowd 1958; Broerse and Zwaan 1966). The importance attached to the beginnings of written words is further emphasised by the convention of arranging dictionaries and other word listings according to a left-to-right alphabetic principles. It may be that this processing habit, developed in the visual modality, is carried over to auditory processing. It is interesting to note that although evidence from lexical selection errors or 'malapropisms' suggests that adults use the beginnings of words as one of the organising principles in their mental lexicons (Fay and Cutler 1977), pre-literate children show no such preference (Aitchison and Straf 1981).

Furthermore, there is evidence that the overt training involved in learning to read engenders an awareness of the phonological structure of a language (Calfee et al 1972; Liberman et al 1974), and that this awareness does not arise spontaneously (Morais et al 1979). While the absence of a conscious awareness of speech as a sequence of phones need not necessarily imply an inability to make use of them in perception and production (English plural formation, for example, is one of many linguistic processes which requires such implicit knowledge, and which preliterate children can perform competently -- see Berko, 1958) it is reasonable to speculate that explicit recognition of such units might facilitate the analysis of the acoustic signal. This is the hypothesis which will be examined in Experiment 1b.

3. Experiment 1b

Experiment 1a established that contextual cues contributed to the process of selection of word candidates, and that subjects could recognise the majority of words before their acoustic offsets (at least if the stimuli were heard in context). However, manipulation of stress pattern had exerted a smaller effect than expected on subjects' performance; although some stress effects were found, these were small in absolute terms and were confined to the earliest stages in the processing of words. No such effects were found when the dependent variables were the isolation point or recognition point. It was hypothesised in the last section that this outcome might be attributable, indirectly, to the undergraduates' exposure to reading, which could, it was suggested, have enhanced their processing skills by increasing their awareness of phonological structure. Therefore, in Experiment 1b, the materials used in Experiment 1a were presented to a group of subjects of similar age and linguistic background to the undergraduates who took part in the earlier experiment. Unlike the undergraduates, however, the second group of subjects spent very little time actively engaged in reading.

In the analyses described in the remainder of the chapter, the data from Experiment 1a have been included since the hypothesis concerns a comparison between the two groups of subjects.

3.1. Method

3.1.1. Materials, Design, Procedure

All aspects of the materials, design and procedure were identical to those described for Experiment 1a.

3.1.2. Subjects

Eighteen subjects, who will be termed the Low Literacy (LL) group to distinguish them from the undergraduate High Literacy (HL) group who took part in Experiment 1a, volunteered to take part in the experiment. All subjects in the Low Literacy group estimated that they spent less than an hour each week actively engaged in reading. The undergraduate estimate, by contrast, had averaged some 15 hours. The mean age of the Low Literacy group was 18.7 years. All were native speakers of Scots English.

3.2. Results

3.2.1. The Phonological Access Hypothesis

As in Experiment 1a, three aspects of processing will be examined: first, subjects' sensitivity to stress levels in the initial portions of words; second, the degree of match between the initial portions of inputs and responses; and finally, cohort entry points.

The strong version of the stressed syllable model predicts first, that subjects can distinguish between stressed and unstressed syllables prior to identifying the words containing them, and second, that only stressed syllables will prove useful cues to the identity of words. A weaker version of this hypothesis will be considered in a later section. By contrast, CM1 predicts that subjects will use the initial syllables of words to gain access to the mental lexicon, regardless of whether the syllables are stressed or unstressed.

3.2.1.1. Sensitivity of subjects to stress pattern of input

In Section 2.2.1.1, it was shown that HL subjects were relatively insensitive to the stress level of initial syllables which they heard. This result is damaging for the strong version of the stressed syllable model because, in order to use stressed syllables to access the mental lexicon, hearers would need to be able to determine, prior to recognising words, which syllables were stressed and which unstressed. In this section, we examine the possibility that another group, the LL subjects, will meet the prerequisite of the model.

For each test word, the gate containing the end of the first vowel was identified as described in Section 2.2.1.1. and the responses offered at that gate were examined to determine the percentage which were stressed on the initial syllable. Monosyllabic words were included in the analysis, and were considered to be initially stressed. For the claims of the Stressed Syllable model to be supported, all (or at least the majority) of responses in the Syll-1 condition would need to be stressed on the first syllable, while this would be the case for very few responses in the Syll-2 condition.

Table 4.21: Percentage of responses with initial stress offered at the gate containing the end of the first vowel (Experiment 1b)

		LITERACY						
		High			Low			
		STRESS						
		Syll-1	Syll-2	(Mean)	Syll-1	Syll-2	(Mean)	Mean
CONTEXT	none	86	70	(78)	87	84	(86)	82
	short	95	33	(64)	94	63	(79)	71
	long	97	11	(54)	89	38	(64)	59
Mean		93	38	(66)	90	62	(76)	

Table 4.22: ANOVA for data in Table 4.21 in relation to Literacy, Stress and Context

Source	F1	F2	MinF′
Literacy	(1,34) 25.18 **	(1,46) 16.55 **	(1,80) 9.99 **
Stress	(1,34) 351.49 **	(1,46) 197.51 **	(1,79) 126.45 **
Context	(2,68) 37.98 **	(2,92) 23.95 **	(2,159) 14.69 **
Stress X Literacy	(1,34) 34.28 **	(1,46) 25.58 **	(1,80) 14.65 **
Stress X Context	(2,68) 84.36 **	(2,92) 38.55 **	(2,152) 26.46 **
Stress X Context X Literacy	(2,68) 3.65 *	(2,92) 4.80 *	n.s.
** p < .01	* p < .05		

In post-hoc comparisons of means by Scheffé test, all comparisons reached significance at p < .01 (by subjects), except for the following: for the Stress X Literacy interaction, the means for Syll-1 words did not differ across the two Literacy groups; for the Stress X Context interaction, the No Context means of 87 for Syll-1 and 77 for Syll-2 were not reliably different.

Contrary to the explanation proposed above, the LL subjects were *less* accurate than the HL group at identifying the stress level of initial syllables: in the Syll-1 condition they offered fewer responses with initial stress, while in the Syll-2 condition they offered more. In the No Context condition, their level of accuracy was only 52% (6% lower than that achieved by the HL subjects).

The pattern of responses is amenable to the same explanation as that provided in Section 2.2.1.1. In the with context conditions, the majority of responses offered were the test word itself, resulting in apparently high accuracy in both Syll-1 and Syll-2 conditions. In the No Context condition, most of the responses in the initially stressed group were monosyllables; of the 94 polysyllabic responses in the LL/Syll-1 condition, 80 were initially stressed, compared with 53 of the 69 responses in the LL/Syll-2 condition. In short, the LL subjects seem no more capable than their HL counterparts of determining the stress level of syllables prior to word recognition.

3.2.1.2. Phonological match between initial portion of input and responses

As in Experiment 1a, the responses produced at the gate containing the end of the first vowel in each test word were collected and analysed with respect to the extent of their match with the input. If the claims of CM1 are correct, the amount of agreement should be high and should be evenly balanced across both Stress categories. The Stressed Syllable model, by contrast, predicts that more agreement should be found in words stressed on the first syllable than in words stressed on the second syllable.

The results of this analysis are presented in Tables 4.23 and 4.24 and in Figure 4.1. Overall, HL subjects were more likely to produce responses which matched the input than were the LL subjects. The Stress effect was significant, but interacted with Literacy such that although the HL and LL means for words stressed on the first syllable did not differ reliably, all other comparisons were significant at the $p < .01$ level (by-subjects Scheffé test). The decrement in performance for initially unstressed words over initially stressed words was more marked for LLs than HLs, consistent with the hypothesis presented above: unstressed syllables caused more difficulties for the LL group than for the HL group.

It was argued above (Section 2.2.1.2) that although stressed syllables were more intelligible than unstressed syllables, the stressed syllable hypothesis, even in its weak form, could not explain the results, because it predicted that hearers would map a syllabic stimulus on to the stressed portion of a lexical entry, but not on to an unstressed syllable; the latter behaviour was, however, observed. A similar analysis was conducted of the Syll-1/no context responses of the LL subjects. Of the 94 polysyllabic responses in this condition, the LLs produced 14 words stressed on the second syllable. Of these, 7 involved the mapping of the stimulus onto unstressed syllables. The phenomenon therefore occurs with similar frequency in both the HL and the LL sample, and argues against the stressed syllable account of phonological access.

Table 4.23: Mean percentage of cases in which subjects produced responses consistent with C(C)V word-initial stimulus (Experiment 1b)

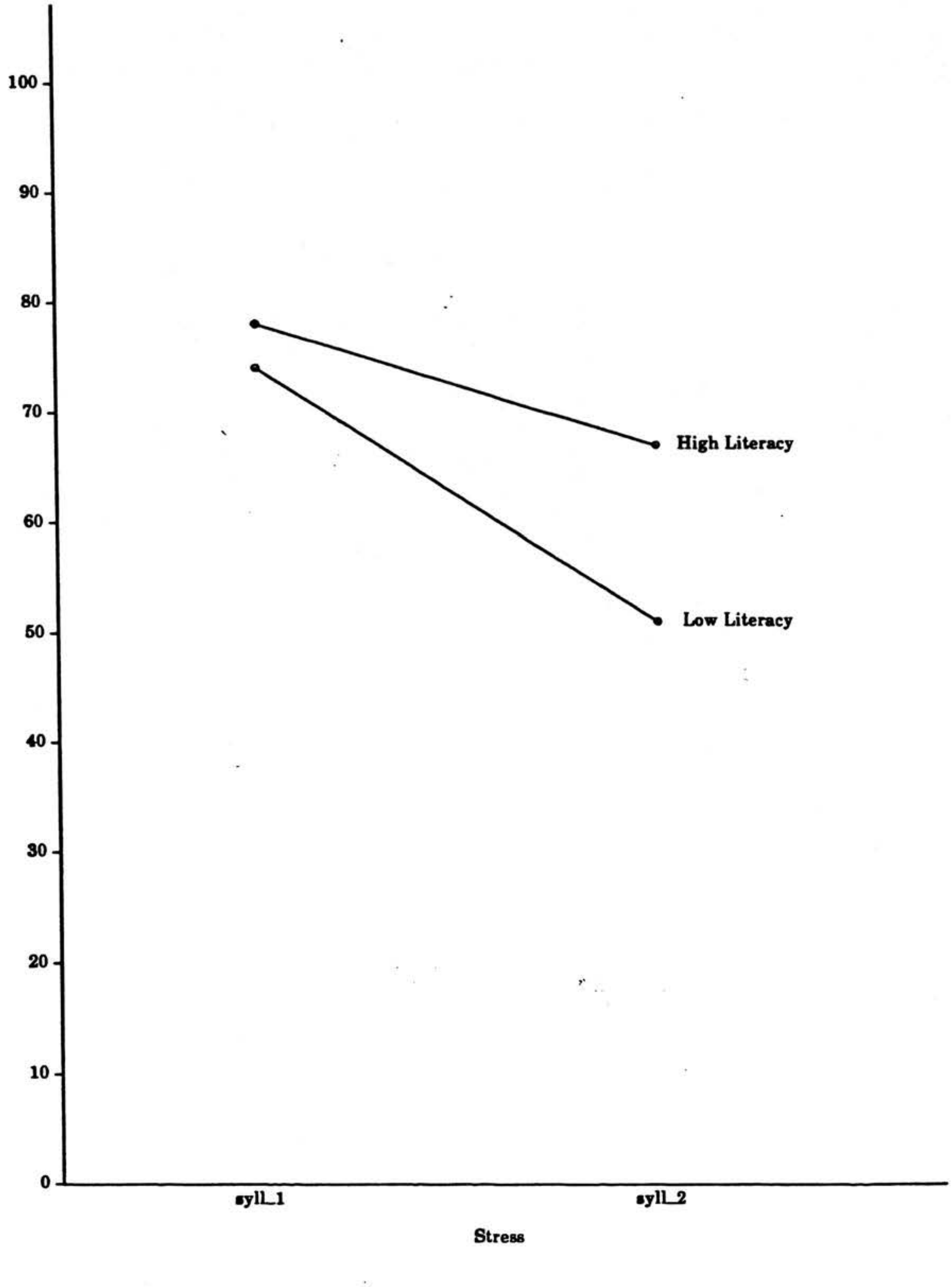
		LITERACY						
		High			Low			
		STRESS						
		Syll-1	Syll-2	(Mean)	Syll-1	Syll-2	(Mean)	Mean
	none	61	48	(55)	63	37	(50)	53
CONTEXT	short	80	67	(74)	77	50	(64)	69
	long	94	85	(90)	81	65	(73)	82
Mean		78	67	(73)	74	51	(63)	

Table 4.24: ANOVA for data in Table 4.23 in relation to Literacy, Stress and Context

Source	F1	F2	MinF′
Literacy	(1,34) 10.11 **	(1,46) 22.32 **	(1,62) 6.96 **
Stress	(1,34) 127.25 **	(1,46) 11.09 **	(1,54) 10.20 **
Context	(2,68) 47.49 **	(2,92) 23.56 **	(2,154) 15.75 **
Stress X Literacy	(1,34) 13.03 **	(1,46) 6.64 *	(1,77) 4.40 *
** p < .01	* p < .05		

The Context effect was significant, with all three means differing at $p < .01$, by-subjects Scheffé test.

Figure 4.1:
Percentage of responses consistent with initial CCV of stimulus, Experiment 1 (Stress X Literacy interaction)



3.2.1.3. Cohort Entry Points

For this analysis, subjects were considered to have entered the appropriate cohort when they began to respond with words matching the target in initial C(C)V, and continued to do so. Cohort Entry points were computed and the resulting data input to by-subjects and by-materials ANOVAs.

Table 4.25: Mean number of 50msec gates heard before Cohort Entry (Experiment 1b)

		LITERACY						
		High			Low			
		STRESS						
		Syll-1	Syll-2	(Mean)	Syll-1	Syll-2	(Mean)	Mean
CONTEXT	none	4.3	4.4	(4.4)	4.7	5.0	(4.9)	4.7
	short	2.9	2.9	(2.9)	3.5	3.9	(3.7)	3.3
	long	2.0	2.0	(2.0)	2.9	2.7	(2.8)	2.4
Mean		3.1	3.1	(3.1)	3.7	3.8	(3.8)	

Table 4.26: ANOVA for data in Table 4.25 in relation to Literacy, Stress and Context

Source	F1	F2	MinF′
Literacy	(1,34) 25.31**	(1,46) 46.04**	(1,67) 16.33**
Stress	n.s.	n.s.	---
Context	(2,68) 216.72**	(2,92) 52.85**	(2,132) 42.49**
** p < .01	* p < .05		

The stress effect was not significant, but (as discussed above, Section 2.2.1.3) this may well be an artefact of the gating method itself, which fails to take account of the differing lengths of the initial syllables in the two stress conditions. HL subjects entered the appropriate cohort sooner than LL subjects. The Context effect was significant: Cohort Entry points were earlier for words heard in context than for words heard in isolation, while the difference between the two with-context conditions was also significant ($p < 0.01$, by subjects, for all comparisons). No interactions were significant.

In the light of the remarks made in the discussion of Experiment 1a, it is unclear whether the significant Literacy effect is to be interpreted as evidence of deficient bottom-up processing skills on the part of LL subjects, or of deficient top-down processing skills, or of both. If the Context result had supported the claims of the Cohort Model (by failing to reach significance) then we should have been justified in attributing LL subjects' later Cohort Entry scores to their comparatively poor ability to process the acoustic/phonetic input. However, as we have seen in Experiment 1a, the Cohort Entry point dependent variable does not reflect only subjects' bottom-up processing activity, as the Cohort Model predicts it should: rather, the effects of context are also involved. Results presented in Section 3.2.3. will demonstrate that LL subjects' lexical search is also influenced by the presence of context. Discussion of this point will be deferred until after further analyses have been presented.

3.2.1.4. Summary of evidence relating to Phonological Access Hypothesis

The Literacy variable was introduced because it was thought that the LL subjects, because of their comparative lack of exposure to reading and its concomitant emphasis on the importance of the beginnings of words, would be likely to experience difficulty in processing unstressed syllables. This appears to be the case: when the dependent variable involved the match between the C(C)V word-initial stimulus and the responses offered, their performance on Syll-1 words was similar to that of the HL subjects, but they suffered a more marked decrement on Syll-2 stimuli.

3.2.2. The Early Recognition Hypothesis

The dependent variables examined to determine whether, as Marslen-Wilson (1987) suggests, the majority of words are recognised by their acoustic offset, are, as in Section 2.2, the percentage of words isolated and recognised by the final gate.

3.2.2.1. Percentage of words successfully isolated by the final gate

As in section 2.2. above, a subject was considered to have isolated a word at the point at which that word begins to be offered as a response and subsequently continues to be offered. The percentage of words in each cell isolated by the final gate was the input to a pair of three-way ANOVAs. Cell means are presented in Table 4.27, and the corresponding ANOVA results in Table 4.28.

Table 4.27: Mean percentage of words successfully isolated by final gate (Experiment 1b)

		LITERACY						
		High			Low			
		STRESS						
		Syll-1	Syll-2	(Mean)	Syll-1	Syll-2	(Mean)	Mean
CONTEXT	none	96	98	(97)	91	90	(91)	94
	short	97	99	(98)	95	99	(97)	98
	long	100	99	(100)	94	100	(97)	99
Mean		97	99	(98)	93	96	(95)	

Table 4.28: ANOVA for data in Table 4.27 in relation to Literacy, Stress and Context

Source	F1	F2	MinF′
Literacy	(1,34) 19.61**	(1,46) 17.71 **	(1,79) 9.05 **
Stress	(1,34) 4.84 *	n.s.	---
Context	(2,68) 7.84 **	(2,92) 7.27 **	(2,157) 3.77 *
** $p < .01$	* $p < .05$		

As in the corresponding analysis in Experiment 1a, the level of successful isolation was high in all conditions. Stress reached significance in the by-subjects analysis only. However, contrary to the prediction being examined, subjects in the LL group, like the HL subjects, isolated words stressed on the second syllable somewhat earlier than words stressed on the first syllable.

Literacy was significant in all analyses, with subjects in the High group isolating more words than subjects in the Low group. The Context effect was also significant: a post-hoc comparison of means by Scheffé test revealed that more words were identified in the two with-context conditions than in the No Context condition ($p<0.01$ by subjects in both cases); the Short and Long Context conditions did not differ. None of the interactions between variables were significant.

3.2.2.2. Percentage of words successfully recognised by final gate

According to Grosjean's (1980) terminology, recognition occurs when a word has been isolated and the subject begins to offer confidence ratings of a high level (in this case, the level selected was 4 on a scale running from one to five). The mean percentages of successful recognitions are displayed in Table 4.29, with the corresponding F-values in Table 4.30.

Table 4.29: Mean percentage of words recognised by the final gate (Experiment 1b)

		LITERACY						
		High			Low			
		STRESS						
		Syll-1	Syll-2	(Mean)	Syll-1	Syll-2	(Mean)	Mean
CONTEXT	none	85	91	(88)	76	86	(81)	85
	short	94	98	(96)	91	92	(92)	94
	long	100	99	(100)	85	97	(91)	96
Mean		93	96	(95)	84	92	(88)	

Table 4.30: ANOVA for data in Table 4.29 in relation to Literacy, Stress and Context

Source	F1	F2	MinF′
Literacy	(1,34) 6.41 *	(1,46) 26.09 **	(1,47) 5.15 *
Stress	(1,34) 15.93 **	n.s.	---
Context	(2,68) 14.25 **	(2,92) 10.74 **	(2,160) 6.12 *
** $p < .01$	* $p < .05$		

The Stress main effect was significant in the by-subjects analysis only. Subjects in the HL group recognised more words than subjects in the LL group. The difference between all three context conditions was significant ($p < .01$, by-subjects Scheffé test). No interactions reached significance.

3.2.2.3. Summary of Evidence relating to the Early Recognition Hypothesis

Like the HL subjects, the LL group were able to recognise most words in context before their acoustic offset; In general, the LL subjects had more difficulty than their HL counterparts in recognising words, but lexical stress pattern had little role to play in this: the stress effects discernible when dependent variables relate to lexical access were no longer apparent at the lexical retrieval stage.

3.2.3. The Context-Free Hypothesis

As in Section 2.2.1, three sets of analyses of the data collected from gates 1, 2 and 3 are presented. The first relates to the grammatical category of responses, the second to their semantic appropriateness, and the third to the size of the word candidate set accessed. In all three cases, the Cohort Model predicts no significant difference between context conditions.

3.2.3.1. Grammatical category of responses

The number of responses with the word class label 'noun' was the dependent variable in the analyses presented in Tables 4.31 and 4.32.

Table 4.31: Mean percentage of responses in gates 1, 2 and 3 with appropriate grammatical category (Experiment 1b)

		LITERACY						
		High			Low			
		STRESS						
		Syll-1	Syll-2	(Mean)	Syll-1	Syll-2	(Mean)	Mean
CONTEXT	none	74	71	(73)	71	74	(73)	73
	short	90	84	(87)	84	80	(82)	84
	long	98	94	(96)	94	87	(91)	93
Mean		85	83	(84)	83	80	(82)	

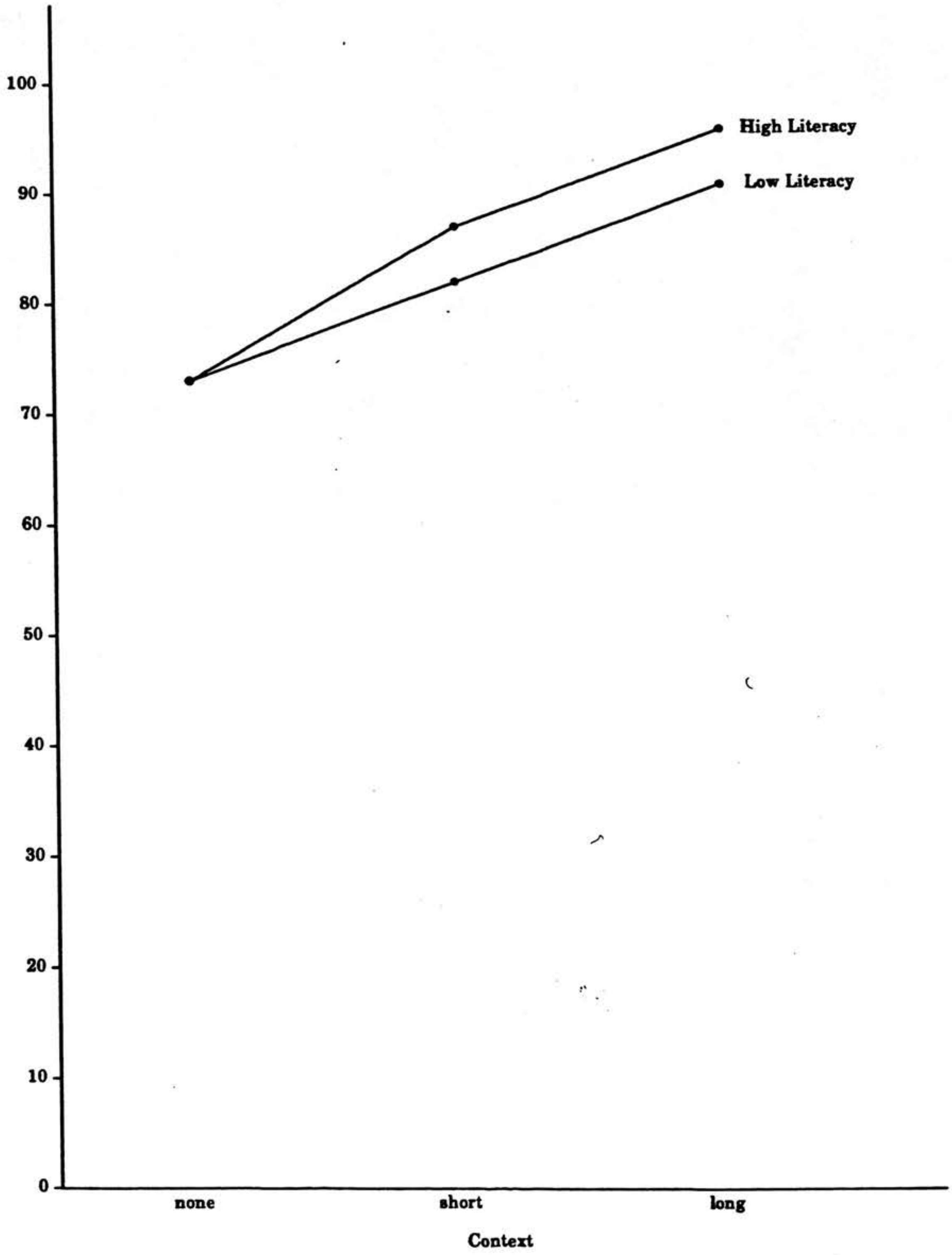
Table 4.32: ANOVA for data in Table 4.31 in relation to Literacy, Stress and Context

Source	F1	F2	MinF′
Literacy	n.s.	(1,46) 8.67 **	---
Stress	(1,34) 4.70 *	n.s.	---
Context	(2,68) 51.19**	(2,92) 13.16 **	(2,133) 10.47 **
Context X Literacy	(2,68) 4.81*	(2,92) 5.43**	(2,153) 2.55*
** $p < .01$	* $p < .05$		

According to the by-materials analysis, HL subjects tended to offer more noun responses than LL subjects did; but neither F1 nor Min F' were significant. Stress, similarly, proved significant in one analysis only, with words stressed on the first syllable producing more nouns than words stressed on the second syllable. It was suggested above (section 2.2.1.2.) that this difference might be attributable to subjects' intuitive awareness that nouns and verbs tend to be stressed on the first and second syllables, respectively.

Context was the only main effect which proved significant in all three analyses. Interestingly, from the point of view of the main hypotheses currently under consideration, a Context by Literacy interaction (illustrated in Figure 4.2) was found: by-subjects Scheffé tests showed that although all three Context conditions differed from each other for the HL subjects, for LL subjects only the presence of the highly constraining Long Context enhanced their likelihood of producing nouns (compared with the No Context control condition). Note, however, that one set of subjects who took part in the original cloze procedure used in the construction of the materials were from the Low Literacy population (see section 2.1.3.); these subjects produced syntactically appropriate completions to the sentences in the vast majority of cases.

Figure 4.2:
Percentage of responses which were nouns,
Experiment 1 (Context X Literacy interaction)



It would seem, therefore, that LL subjects have problems in making use of syntactic cues during the auditory (though not visual) processing of some sentences. Time pressures alone could not account for this difference between the cloze test result and the gating result: in the gating experiment, subjects had up to 16 seconds in which to respond.

In spite of the Context X Literacy interaction just described, both groups of subjects appear to be using syntactic context to limit the size of the word initial cohort. Tyler (personal communication) has questioned the assumption, implicit in the description of the analysis just presented, that subjects apprehended the information that the test word occurred sentence-finally, and has suggested that a more appropriate dependent variable might be the number of responses whose word classes are permitted to follow determiners, without necessarily completing a sentence. For example, a determiner may be followed by an adverb (e.g. *a hastily prepared meal*), although the sequence [determiner + adverb] may not terminate a sentence (e.g. **He ate a hastily.*). Accordingly, the non-noun responses offered by subjects were examined, to see whether the inclusion of responses in the categories which can follow determiners might alter the experimental finding.

Of the non-noun responses offered in the with-context conditions, the majority (191 out of 202) became appropriate responses under this new criterion. In most cases where subjects offered responses which were still inappropriate (e.g. pronouns), this appeared to result from a mis-segmentation of the preceding sentence fragment (see Section 3.3. for discussion of this point). Given the already high level of noun responses offered in the with-context conditions, the interpretation offered above, that hearers were using contextual information to delimit the word-initial cohort, is further strengthened. By contrast, only 32 of the 220 non-noun responses offered in the no-context condition migrated into the 'appropriate' group when the new criterion was adopted. The majority of the non-noun responses in this condition were function words (prepositions, conjunctions, determiners, etc) as well as a few verbs.

Since the level of noun responses which subjects produced in the No Context condition was markedly lower than that observed in the other conditions, the addition of these few extra responses would not decrease the differential between it and the with-context conditions, and the interpretation offered above can be sustained.

3.2.3.2. Semantic appropriateness of responses

Semantic appropriateness of responses was assessed in the manner described for Experiment 1a: for each of the two comparisons (No Context versus Short Context, and No Context versus Long Context) a separate pair of judges was asked to evaluate the appropriateness of each response made in the first three gates of the No Context and the with context condition against the sentence introduction heard in the with context condition. That is, for the first analysis (section 3.2.1.2.1.) No Context responses and Short Context responses were evaluated against the Short Context sentence, while in the second analysis (section 3.2.1.2.2.) No Context and Long Context responses were evaluated against the Long Context sentences.

3.2.3.2.1. Suitability of responses made in No Context and Short Context conditions as completions of Short Context sentences

Responses offered by subjects in the No Context condition were compared with those produced in the Short Context condition for the next pair of ANOVAs. Table 4.33. shows cell means, Table 4.34, results.

Table 4.33: Percentage of responses offered in the No Context and Short Context conditions which were judged to be appropriate completions of the Short Context sentence (Experiment 1b)

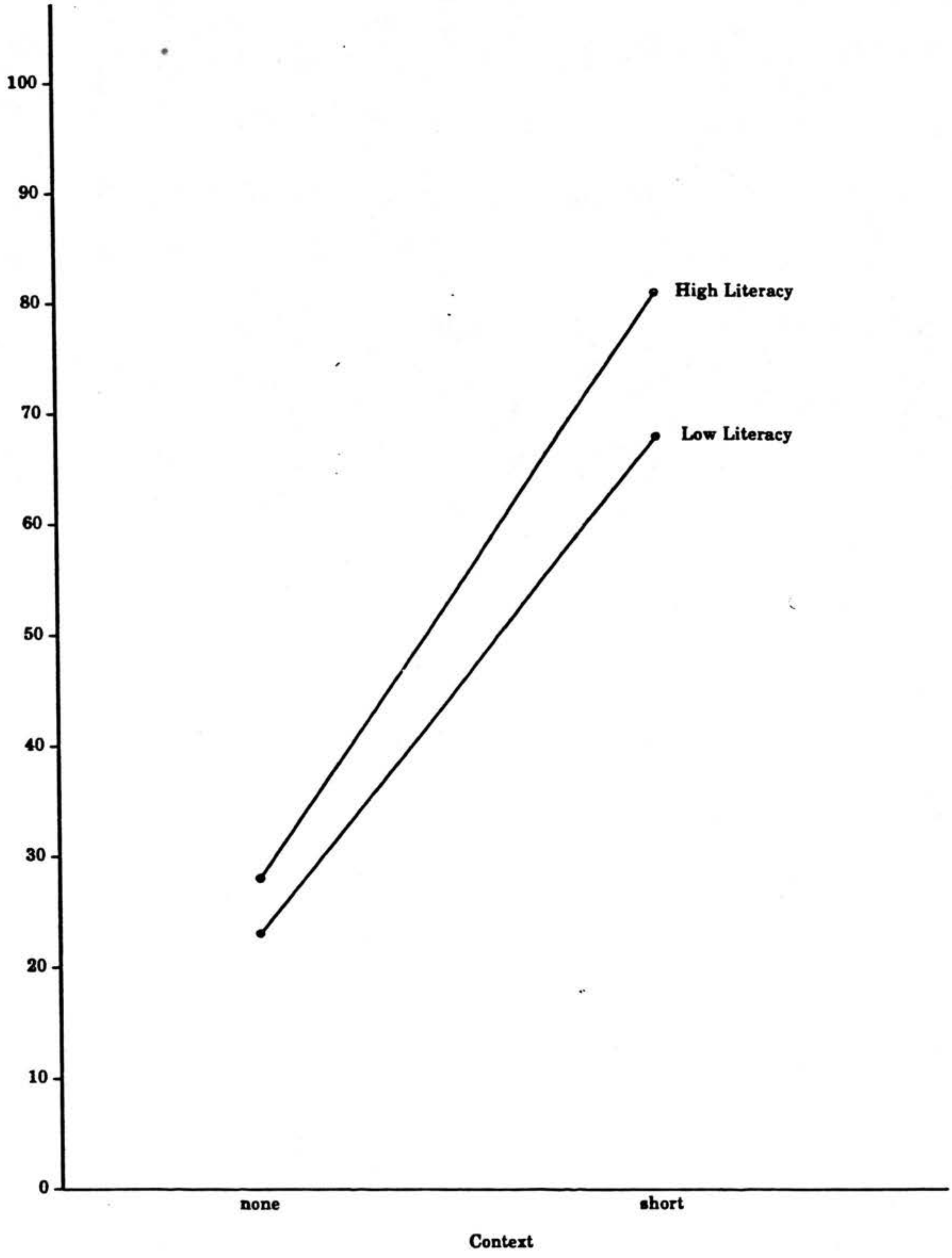
		LITERACY						
		High			Low			
		STRESS						
		Syll-1	Syll-2	(Mean)	Syll-1	Syll-2	(Mean)	Mean
CONTEXT	none	31	24	(28)	24	22	(23)	26
	short	83	79	(81)	72	64	(68)	75
Mean		57	52	(55)	48	43	(46)	

Table 4.34: ANOVA for data in Table 4.33 in relation to Literacy, Stress and Context

Source	F1	F2	MinF′
Literacy	(1,34) 14.66 **	(1,46) 11.72 **	(1,80) 6.51*
Stress	(1,34) 4.85 *	n.s.	---
Context	(2,68) 619.48 **	(2,92) 165.24 **	(1,67) 130.45 **
Context X Literacy	(2,68) 4.51 *	n.s.	---
** p < .01	* p < .05		

HL subjects produced a greater percentage of appropriate responses than LL subjects. Although the Stress effect reached a weak level of significance in the by-subjects analysis, it failed to do so by materials. The Context effect was once again strong: the responses offered in the No Context condition were far less likely to be appropriate completions of the Short Context sentences than the responses produced in the Short Context condition itself. According to the by-subjects analysis, Literacy and Context interacted, as shown in Figure 4.3. Post-hoc comparisons of means by Scheffé test showed that the No Context means did not differ for the two groups of subjects (which is to be expected if the No Context data did indeed constitute a randomly generated set of words), but that the difference between their Short Context means was highly significant ($p < .001$, by-subjects); LL subjects used contextual cues less efficiently than HL subjects. Within Literacy groups, means differed at $p < .001$.

Figure 4.3:
Percentage of responses which were semantically appropriate, Experiment 1 (NC/SC comparison, Context X Literacy interaction)



3.2.3.2.2. Suitability of responses made in No Context and Long Context conditions as completions of Long Context sentences

The same procedure as in the previous section was followed to assess the suitability of responses wade in the Long Context condition. Results of this analysis are shown in Tables 4.35 and 4.36.

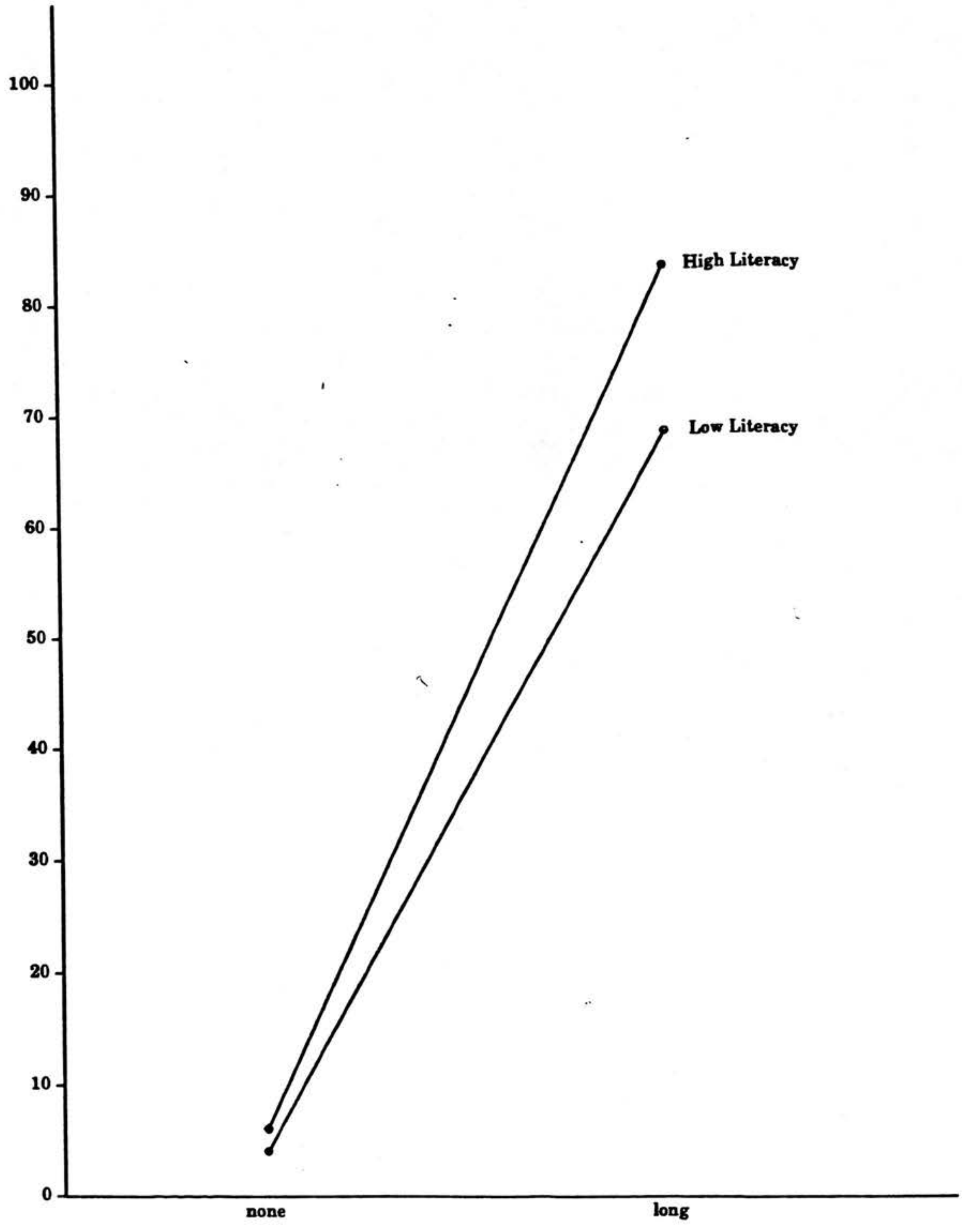
Table 4.35: Percentage responses offered in the No Context and Long Context conditions which were judged appropriate completions of the Long Context sentence (Experiment 1b)

		LITERACY						
		High			Low			
		STRESS						
		Syll-1	Syll-2	(Mean)	Syll-1	Syll-2	(Mean)	Mean
CONTEXT	none	5	7	(6)	6	2	(4)	5
	long	82	86	(84)	68	70	(69)	77
Mean		44	47	(46)	37	36	(37)	

Table 4.36: ANOVA for data in Table 4.35 in relation to Literacy, Stress and Context

Source	F1	F2	MinF'
Literacy	(1,34) 19.93 **	(1,46) 28.46 **	(1,73) 11.72 **
Stress	n.s.	n.s.	---
Context	(1,34) 1139.90 **	(1,46) 666.88 **	(1,79) 420.74 **
Context X Literacy	(1,34) 10.06 **	(1,46) 15.04 **	(1,71) 6.03 *
** $p < .01$	* $p < .05$		

Figure 4.4:
Percentage of responses which were semantically appropriate,
Experiment 1 (NC/LC comparison, Context X Literacy interaction)



The Literacy effect was significant: HL subjects tended to offer more appropriate responses than LL subjects. No significant differences were found between levels of the Stress variable. The difference between the two Context conditions was even more marked than in the previous analysis: stimuli heard in the Long Context condition elicited many more appropriate responses than those heard in the No Context condition. The Context by Literacy interaction (see Figure 4.4) which was found in the by-subjects analysis in the previous section now reached strong significance for all three values of F; post-hoc comparisons of means by Scheffé test showed that the No Context means did not differ for the two groups of subjects, but that the difference between their Long Context means was highly significant ($p < .001$, by-subjects). Once again, HL subjects were more efficient in their use of contextual cues than their LL counterparts.

3.2.3.3. Type-token ratio of responses

The Cohort Model predicts that the number of distinct word candidates accessed will remain constant, regardless of the presence or absence of context. In terms of gating responses, then, the type-token ration should not vary between conditions. Type-token ratios for responses to each word were computed as described in section 2.2.1.3. These scores were input to a single, by-materials ANOVA. The means for each condition are shown in Table 4.28, and the results of the analysis in Table 4.29.

Table 4.37: Type-token ratio of responses offered during gates 1, 2 and 3 (Experiment 1b)

		LITERACY						
		High			Low			
		STRESS						
		Syll-1	Syll-2	(Mean)	Syll-1	Syll-2	(Mean)	Mean
	none	.73	.65	(.69)	.75	.66	(.71)	.70
CONTEXT	short	.47	.46	(.47)	.50	.51	(.51)	.49
	long	.26	.25	(.26)	.39	.36	(.38)	.32
Mean		.49	.45	(.47)	.55	.51	(.53)	

Table 4.38: ANOVA for data in Table 4.37 in relation to Literacy, Stress and Context

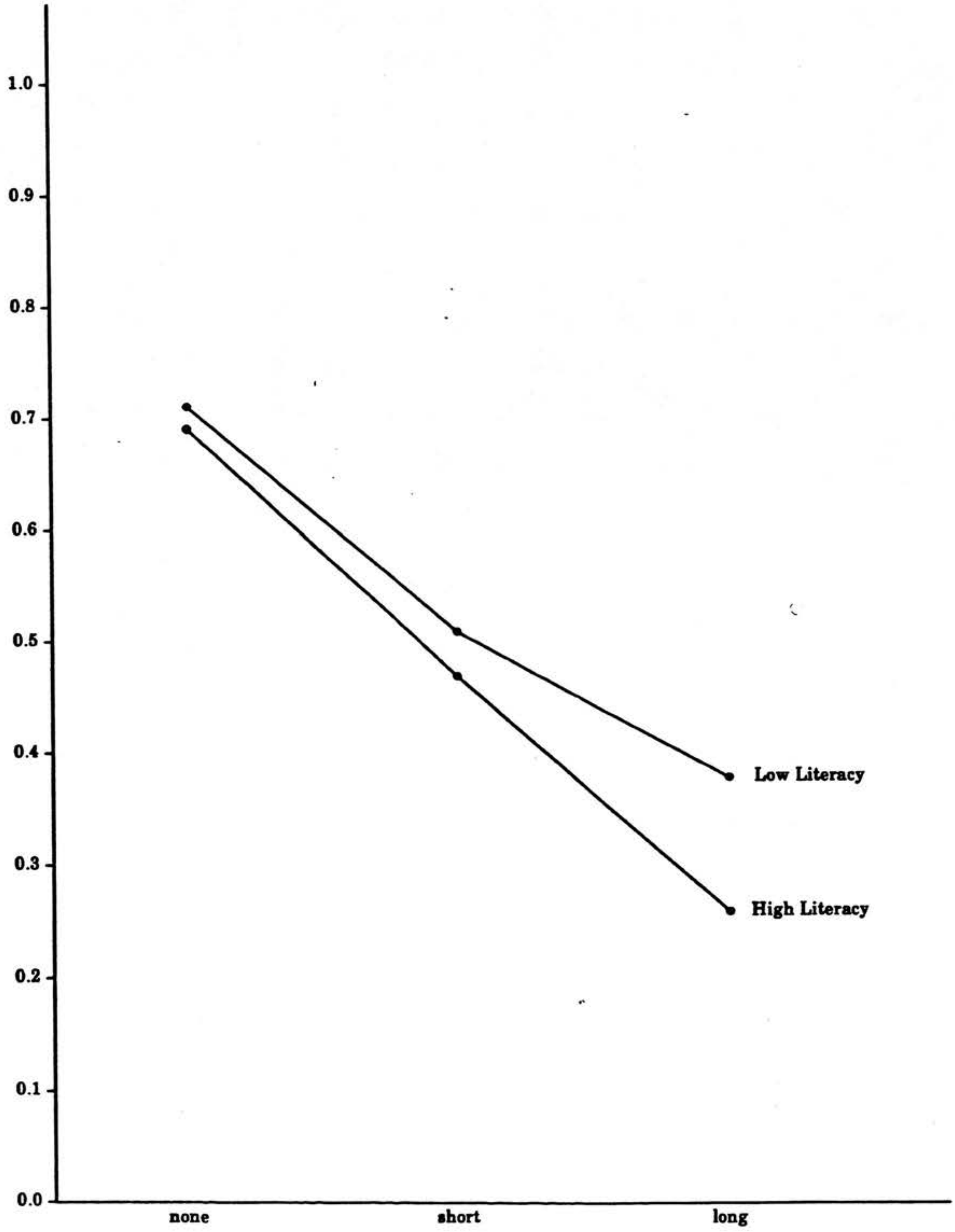
Source	F
Literacy	(1,46) 14.93 **
Stress	n.s.
Context	(2,92) 137.15 **
Context X Literacy	(2,92) 4.33*
** p < .01	* p < .05

As in the other analyses, the Stress variable was not significant. Subjects in the HL group offered fewer different responses per word than subjects in the LL group. Context acted to limit the number of words accessed during Cohort initiation; according to by-subjects Scheffé tests, all three means differed at $p < .01$. Once more, a Context by Literacy interaction (see Figure 4.5) was observed; although, within each Literacy group, all three Context means differed, across Literacy groups, the main difference was between the Long Context conditions ($p < 0.01$, Scheffé test).

3.2.3.4. Summary of evidence relating to the Context-Free Hypothesis

As in Experiment 1a, hearers appear to use context even in the earliest stages of processing to constrain their choice of word candidates. For dependent variables relating to syntactic and semantic processing and to the number of distinct word candidates accessed, effects of context are discernible in the responses of both groups of subjects. However, the LLs' performance was not identical in all respects to that of the HLs. Although both groups of subjects used context in the way described, HLs did so more efficiently than LLs: for the latter group, only the presence of the highly constraining Long Context sentence introduction enhanced their performance.

Figure 4.5:
Type-Token ratio of responses,
Experiment 1 (Context X Literacy interaction)



3.2.4. Summary and Discussion of Experiments 1a and 1b

The results of Experiments 1a and 1b are summarised in Table 4.39. Experiment 1b set out to answer the question of whether the apparent lack of difficulty experienced by subjects in Experiment 1a was due to their highly-developed bottom-up processing skills. There is some support for this explanation in the results of Experiment 1b. While LL subjects produced similar results in the Syll-1 condition, the Syll-2 condition caused them comparatively more difficulty.

Table 4.39: results of Experiments 1a and 1b

Note:

- the symbol * indicates analyses significant for F1, F2 and MinF'
- analyses 1 - 3, phonological access hypothesis; 4 and 5, early recognition hypothesis; 6 - 9, context-free hypothesis.

Dependent Variable	Literacy	Stress	Context	Interaction
1. Percentage of responses with appropriate stress level in initial syllable	*	*	*	Stress X Literacy; Stress X Context
2. Percentage of responses which matched test word in initial C(C)V	*	*	*	Stress X Literacy
3. Cohort Entry Points	*		*	
4. Percentage of successful isolations by final gate	*	F1	*	
5. Percentage of successful recognitions by final gate	*	F1	*	
6. Grammatical category of responses 1, 2 & 3	F2	F1	*	Context X Literacy
7. Semantic appropriateness of no-context vs short-context responses	*	F1	*	Context X Literacy
8. Semantic appropriateness of no-context vs long-context responses	*		*	Context X Literacy
9. Type/token ratio of responses	*		*	Context X Literacy

On the other hand, when the dependent variables examined related to retrieval, stress effects were not found. Thus, just like the HL subjects, the LLs were able to recover from the detrimental influence of unstressed initial syllables by the later stages of processing.

Differences between the two groups of subjects emerged in their use of context: although both LL and HL subjects used syntactic and semantic context to constrain candidate selection during the lexical access phase, LL subjects were less skilled at doing so than HL subjects. Note, however, that this difference between the two Literacy groups was not apparent when the materials were pretested by visual cloze procedure; thus, it appears that the problem for LL subjects is one specific to the auditory modality. It was pointed out above (section 2.3.) that the ability to use contextual cues in auditory word recognition presupposes the ability to process the acoustic input representing that context, and to this extent, the issue is one of bottom-up processing ability. However, any attempt to resolve this question would lead to investigations which it would be inappropriate to pursue here.

The phonological access and early recognition hypotheses will be examined further in Experiment 2. On the basis of the results presented in this chapter, however, it seems that we can reject the context-free hypothesis: contrary to the claims of the Cohort Model, subjects do make use of contextual information when accessing a set of word candidates. However, if we accept that hearers *can* make use of contextual information in a way which results in fewer word candidates being passed on to the lexical retrieval stage, we must ask why they sometimes fail to do so. It has been noted above that the utilisation of left context presupposes the accurate recognition of the words of which that context is composed. This is not a trivial consideration. Even in Experiment 1, which used carefully articulated materials, there is evidence that subjects may misinterpret the auditory input. For example, most of the subjects who heard the sentence fragment

(4.4) When he shows us his tricks, we all admire a ...

followed by the initial gate of the stimulus *magician* (which they correctly identified as /m/) appeared to have interpreted the sequence as

(4.5) When he shows us his tricks, we all admire him ...

since they offered responses consistent with such a parse (e.g. *immensely, a lot*). Not a single subject parsed the phonetic sequence appropriately, despite the fact that all the test sentences have a similar syntactic structure, with a determiner (*a* or *the*) preceding the target. The same phenomenon was observed in the Short Context condition. This erroneous interpretation persisted for several gates in the case of most subjects. Similarly, the sentence introductions in 4.6. and 4.8. (and their corresponding Short Context introductions) seem to have been interpreted by most subjects as those in 4.7. and 4.9:

(4.6) When John threw his racquet in the air, it hit a ... (test word *spectator*)
(4.7) When John threw his racquet in the air, it hit us ... (example of response: *hard*)

(4.8) Although he sometimes watches our team, Bill is not a ... (test word *supporter*)
(4.9) Although he sometimes watches our team, Bill is not as ... (example of response: *clever*).

Thus, although hearers make use of contextual cues as they perceive them, their perceptions are not always correct, even when the speech input is very carefully articulated. We will return to this issue in later chapters, when experiments using conversational materials are described.

As far as the phonological access hypothesis is concerned, full support is found neither for the claims of the Stressed Syllable model nor for the predictions of CM1. Some of the results of Experiment 1 have indicated that unstressed syllables cause difficulties at the lexical access stage, although these results were less convincing than had been anticipated. It may be the case that, in carefully articulated speech of the kind produced by the speaker in Experiment 1, the decrement in clarity in initial unstressed as compared with stressed syllables is not sufficient to have a serious effect on hearers' processing. However, in conversational speech, phenomena such as reduction are far more common than they are in read speech (Brown, 1977; Shockey, forthcoming). Thus one might hypothesise that the invariance of stressed syllables might prove increasingly useful to hearers as the quality of the acoustic input in general declines. This hypothesis is tested in the experiments described in the next chapter.

Chapter Notes

(1) Although the versions of the Cohort Model detailed in the 1980 and 1987 papers differ in certain respects, discussed in Chapter 2, both claim that it is the initial portions of words which are used to access lexical representations.

(2) A model involving full vowels has been proposed by Cutler and Norris (1988). However, they claim that input is mapped onto *initial* syllables, rather than onto any syllable containing a full vowel in the lexical entry; their model consequently has more predictive power than the one hypothesised here.

(3) In fact, several pairs matched for their initial CVC.

(4) One group of judges was drawn from the undergraduate population of Edinburgh University. The other was drawn from the same 'Low Literacy' population as the subjects in Experiment 1b.

(5) The onset of a test word was considered to be the onset of the closure phase if the initial segment was a stop and of frication if the initial segment was a fricative. In two of the twenty-four pairs, the word-initial segment was a nasal, a segment whose onset is more difficult to locate than that of a stop or a fricative. Indeed, it is for their supposed immunity to co-articulatory effects that other researchers have selected as their test materials only words beginning with stops or fricatives (note, however, that this belief is ill-founded: stop consonants, in particular, are characterised by an anticipatory effect on the formants of the vowel preceding them: see Fry, 1979). While it would have been preferable to use only words whose initial segments belonged to one of these two classes, the constraints outlined earlier had limited the number of words available, and necessitated the inclusion of the two pairs with initial nasals. Although it might have been possible, using the auditory signal, to locate the onset of nasalisation in the preceding vowel and to mark that as the onset of the test word, the stimulus which would have resulted in the NC condition would have been perceived as beginning with a vowel and would therefore have been misleading for subjects. The solution which was adopted was therefore to move the cursor frame by frame until the

vowel which preceded the nasal could no longer be detected when the test word was presented in isolation and to adopt that point as the onset of the word. Although such a procedure must necessarily have introduced some anticipatory co-articulation effects they would be duplicated across the two stress conditions, since both members of each pair began with the same consonant.

(6) I am grateful to Norman Dryden of the Department of Linguistics for writing the program which performed the gating of the stimuli in Experiments 1, 2 and 3.

(7) As Section 2.2.3. will demonstrate, hearers use context in the word recognition process even from the earliest gates.

(8) The value of 4 rather than 5 was adopted because a few subjects appeared to be reluctant to use the top of the scale, never registering a confidence-rating of 5. It seems unreasonable to suppose that these subjects were *never* fully confident in their responses.

(9) Tyler (personal communication) has pointed out that it is possible that subjects failed to note that the test word was the final word in the sentence, and that a more appropriate dependent variable might therefore be the number of words which fall into grammatical categories which are permitted to follow determiners. See Section 3.2.3.1. for a discussion of this possibility.

CHAPTER FIVE

Experiments using spontaneous speech

1. Introduction

In Chapter 4, the results of Experiment 1 were discussed in relation to three claims of the Cohort Model and of other theories of auditory word recognition. The context-free hypothesis related to the Cohort Model's prediction that during the lexical access phase of processing no effects of context would be discernible; this hypothesis was rejected because of the statistically significant difference between the no-context and with-context conditions in terms of contextual appropriateness of gating responses and because of other differences between the context conditions. The early recognition hypothesis, which stated that polysyllabic words, and all words heard in context, should be recognised by their acoustic offset, received some support: testwords were identified with a high level of accuracy whether they were heard in isolation or in context. However, the evidence relating to the third issue, phonological access, was equivocal: while the stress variable exerted an effect on subjects' ability to process the initial portions of words which was inconsistent with the claims of the 1980 version of the Cohort Model (CM1), unstressed syllables did not cause the processing difficulties which the literature on reduction, discussed in Chapter 2, would lead one to expect.

The materials used in Experiment 1 were produced, like the majority of materials in word recognition experiments, by having a speaker read carefully constructed experimental sentences. While this approach has the advantage of enabling control of relevant variables, it may be that it obscures the very object of interest, namely, fluent speech. Shockey's (1983) study of the operation of phonological rules in read and spontaneous speech showed these two speech styles differed in a variety of ways. Other writers (e.g. Pollack and Pickett 1963; Wheeldon 1985; Bard et al 1988) have reported that subjects frequently fail to identify words in conversational speech. The near-perfect intelligibility of the testwords in Experiment 1, even when they were heard excised from context, demonstrates their dissimilarity to the word tokens produced in casual speech. Perhaps the unusual clarity of the materials in Experiment 1, as in other psycholinguistic studies, has either allowed subjects to adopt a processing strategy not available to them in the speech to which they are usually exposed, or has provided a biassed picture of the processing strategies which hearers habitually adopt. In spontaneous speech, phonological processes such as assimilation and reduction apply much more frequently than in speech produced under formal laboratory conditions. Under such circumstances, the value of stressed syllables as islands of reliability, which subjects can use to initiate lexical access, is likely to increase, as the informativeness of other, unstressed syllables, decreases. In the experiments reported in this chapter, speech style is manipulated in order to test the hypothesis that as the intelligibility of the speech sample decreases (for example, in casual speech utterances) behaviour consistent with the stressed syllable strategy will appear. That is, in conversational speech, access and retrieval should be faster for Syll-1 words than for Syll-2 words.

In Experiments 2 and 3, therefore, the phonological access hypothesis will continue to be examined, but it will be assessed with respect to conversational as well as read speech. However, it is likely that the manipulation of speech style will also shed light on the early recognition issue. The results of Experiment 1 lent support to the Cohort Model's claim that the majority of polysyllabic words would be recognised by acoustic offset. Bard et al.(1988) demonstrated that a substantial percentage of words -- between 19% and 21% -- remain unrecognised by their acoustic offsets in conversational speech, and claimed that this result militated for a rejection of the notion that speech recognition proceeded in a word-by-word, left-to-right manner. It might be argued, however, that activation-based models could in fact account for the kinds of late recognition which Bard et al.report. It was pointed out in Chapter 2 that the more recent version of the Cohort Model allows for some words to be recognised late; specifically, Marslen-Wilson (1987:75) accepts that

> under certain conditions of temporary ambiguity, ... "late" selection will occur, where a word is not only not recognised early, but may not even be identified until the word following it has been heard.

It is possible that the nature of the stimuli in Bard et al.'s experiment resulted in speech which contained many examples of such temporary ambiguity: tokens which, though distinct at the phonological level, nevertheless were homophonous at the phonetic level. This might have arisen if, for example, the majority of the words which Bard et al.included in their experimental materials were monosyllabic (1). Ambiguity of this kind is common in monosyllabic functors: for example, the sequence [@ z] may represent *us, has, is* or *as*. Similarly, two phonologically distinct contentives such as *hand* and *ham* may become ambiguous at the phonetic level in a sequence such as [hambag] (*ham bag* or *hand bag*). However, it would not to be possible to account for late recognition of polysyllabic words in this way, because it is far less likely that ambiguity will persist past acoustic offset in items of this length. This possibility will be examined in the experiments reported in this chapter: the gating responses to polysyllabic words will be analysed to determine the point at which the tokens are identified. If hearers still fail to recognise a substantial percentage of polysyllabic words, we can conclude that the effect cannot be attributable to conditions of temporary ambiguity. This issue will be examined in Experiment 3.

The aims of Experiment 2 were first, to examine further the phonological access hypothesis, and second, to quantify the intelligibility of a set of word tokens in order to examine further the processing of a subset of these utterances. A large number of words excised from the casual speech of six individuals were presented in isolation to subjects in a gating experiment. The stress pattern of the words which were selected was systematically varied. Read tokens of the same words, produced in identical sentential contexts, were also presented. The analyses presented in the next section show that both stress pattern and speech style affect the processing of words, and that these effects were discernible during both the lexical access and lexical retrieval stages.

2. Experiment 2

2.1. Method

2.1.1. Materials

Six female speakers were recorded in three separate conversations. No attempt was made to control for speaker's regional accent although all six were native speakers of English. Of the speakers, three came from Edinburgh, one from Liverpool, one from Newcastle-under-Lyme, and one from Gloucestershire. All six speakers appeared completely relaxed in the setting of the recording studio; they were naive as to the exact purpose of the recording, having merely been told that the research was concerned with some aspect of speech.

A word-level transcription was made of each of the three conversations and, for each speaker, forty-eight polysyllabic content words were randomly selected: polysyllabic contentives with appropriate stress patterns were numbered on the transcript and a random number table was used to select the test items. A word selected in this manner was included in the sample providing it was not interrupted by noise or by the speech of another individual and had not been subject to rhythmically-induced stress shift. If a word did not meet these criteria, it was rejected and the next appropriate word in the transcript was selected until all forty-eight words had been selected for a given speaker.

Of these forty-eight words, twenty-four were stressed on the first syllable, and twenty-four were stressed on the second syllable. In an attempt to match number of syllables across the two Stress classes, twelve words stressed on the first syllable were randomly selected; then the first twelve randomly selected words stressed on the second syllable which matched these words for number of syllables were also selected. This procedure was repeated with words stressed on the second syllable selected before words stressed on the first syllable. This resulted in equal numbers of 2-syllable words in each stress category, for all speakers except Speaker 2: because of the small number of usable words stressed on the second syllable in the conversation of this speaker, 13 words stressed on the first syllable and 10 words stressed on the second syllable were bisyllabic. These words, excised from context, formed the Spontaneous set. The words are presented in Appendix E. Word frequency did not differ across stress categories or across speakers (see Tables 5.1 and 5.2).

The words in their original sentence contexts (2) were then re-recorded a few days later by their original speakers, who were asked to read them aloud. These words, excised from context, formed the Read set. Thus, for each speaker, forty eight words (types) were available, and for each type there were two tokens, making a total of 96 tokens per speaker, or 576 stimuli in all. A computer program was used to generate the stimuli, gated in 50msec increments, in the order required by the design (see below).

Lengths in milliseconds of test words were input to a three-way analysis of variance. Cell means are presented in Table 5.3, with the results of the analysis in Table 5.4. For all speakers, words stressed on the first syllable were shorter than words stressed on the second syllable; this result has been reported elsewhere in the literature (Cutler and Clifton 1984). The main effect of Speaker was significant, with words spoken by Speaker 6 shorter than those of Speakers 3 and 4 ($p<0.01$, Scheffé test by subjects) and Speakers 1 and 5 ($p<0.05$); Speaker 2's words were shorter than those of Speaker 4 only ($p<0.05$). Words excised from conversations were shorter than words excised from read speech, but this variable interacted with Speaker: only for Speaker 1 were Read words reliably longer than Spontaneous.

Table 5.1: Frequencies of test words, Experiment 2

STRESSED SYLLABLE:	First	Second	Mean
SPEAKER			
1	113	104	109
2	108	117	113
3	91	102	97
4	83	92	88
5	141	119	130
6	97	71	84
Mean	105	101	

Table 5.2: ANOVA for data in Table 5.1.

Source	F
Stress	(1,138) 0.36 n.s.
Speaker	(5,138) 0.47 n.s.
Speaker X Stress	((5,138) 0.81 n.s.

Table 5.3: Lengths of test words, Experiment 2

VERSION	Read			Spontaneous		
STRESS	First	Second	(Mean)	First	Second	(Mean)
SPEAKER						
1	471	524	(498)	363	399	(381)
2	412	487	(450)	343	425	(384)
3	480	532	(506)	428	460	(444)
4	493	500	(497)	475	509	(492)
5	416	473	(445)	403	462	(433)
6	357	362	(360)	344	355	(350)
Mean	438	480	(459)	393	435	(414)

Table 5.4: ANOVA for data in Table 5.3.

Source	F
Version	(1,23) 56.52 **
Stress	(1,23) 10.51 **
Speaker	(5,115) 9.50 **
Version X Speaker	(5,115) 9.13 **
** $p < 0.01$	* $p < 0.05$

2.1.2. Design

The design was as follows: Version (2 levels: Read versus Spontaneous) X Stress Assignment (2 levels: First versus Second Syllable) X Speaker (6 levels). Version, Stress Assignment and Speaker were crossed with each other. Subjects were nested in groups, defined according to the variables just described, such that they heard 16 words from each speaker: 4 read words stressed on the first syllable, 4 read words stressed on the second syllable, 4 spontaneous words stressed on the first syllable and 4 spontaneous words stressed on the second syllable. Each word type was heard in one version only by any one subject. Subjects heard the gated stimuli blocked according to Speaker, with an orientation sentence preceding the block. The orientation sentences are listed in Appendix F; these transcripts were also available to the subjects. Within each speaker block, stimuli were presented in random order. Between gates subjects had 4 seconds in which to respond. The first of a new series of gates (i.e. a new stimulus) was preceded by a warning tone; a different tone was repeated shortly before the presentation of each gate. Six tapes, conforming to all of these requirements, were constructed. The order of presentation of the Speaker-defined blocks was varied across the 6 tapes, such that all six speakers were heard in all possible positions.

2.1.3. Subjects

Twenty-four subjects, students and staff of the Department of Linguistics, took part in the experiment. All were native speakers of English.

2.1.4. Procedure

At the start of the session, which lasted approximately 45 minutes, subjects heard taped instructions and three practice trials (which used stimuli spoken by an individual not represented in the main experiment). The practice trials were not varied across subjects.

Subjects heard the tape in a soundproof booth. They were instructed to respond aloud after each and every presentation of a word or part of a word. Their responses were recorded on cassette and from this recording the transcription of their responses made at the time of the experiment was later checked.

2.2. Results of Experiment 2

For each of the analyses reported, two values of F were computed to allow the calculation of MinF′ (Clark 1973). The Speaker variable yielded results which are uninteresting in terms of the hypotheses currently under investigation: while differences in overall intelligibility were found between speakers, the pattern of results as far as the Stress and Version variables were concerned was consistent. Therefore, in order to improve the comprehensibility of the results tables in the text, in most cases speaker means were collapsed and only the four Stress X Version means are reported.

For all ANOVA tables presented in this chapter, main effects and interactions which reach significance for neither F1 nor F2 are omitted.

2.2.1. Intelligibility of test words

One aim of Experiment 2 was to identify a set of stimuli which would be the subject of further analysis and experimentation. Accordingly, subjects' ability to identify stimuli presented in isolation was of interest. The first dependent variable examined was the percentage of successful identifications of test items by their final gate (see Tables 5.5 and 5.6). The 576 stimuli used in this experiment were each presented to four subjects, yielding a total of 2304 recognition outcomes to be analysed.

More read words than spontaneous, and more words stressed on the first syllable than words stressed on the second syllable, were isolated by the final gate. Version interacted with Speaker; all differences were significant at $p < 0.01$ (Scheffé test by subjects) except for the comparison between the Read means of Speaker 1 and Speaker 5, which was not significant.

Table 5.5: Percentage of trials in Experiment 2 in which the test word was correctly identified by the final gate (raw scores in brackets)

VERSION:	Read	Spontaneous	Mean
STRESS:			
Syll1	83 (478)	58 (334)	71 (406)
Syll2	69 (397)	42 (241)	56 (319)
Mean	76 (438)	50 (288)	

Table 5.6: ANOVA for data in Table 5.5, related to Stress, Version and Speaker

Source	F1	F2	Min F'
Version	(1,23)111.01**	(1,276)103.84**	(1,90) 53.65**
Stress	(1,23) 34.60**	(1,276) 25.95**	(1,109)14.83**
Speaker	(5,115) 7.00**	(5,276) 2.58*	(5,52) 1.89 n.s.
Version X Speaker	(5,115) 7.55**	(5,276) 4.05**	(5,385) 2.64*
Version X Stress X Speaker	(5,115) 3.97**	n.s.	-----
**p<0.01	*p<0.05		

The level of intelligibility of the materials used in Experiment 2 is markedly higher for both read and spontaneous tokens than that reported by Wheeldon (1985), who also presented excised tokens in isolation for identification by hearers. Two differences between the present experimental materials and those of Wheeldon are implicated in this discrepancy. First, Wheeldon made no distinction between content and function words in her analysis; second, since only polysyllabic materials were selected, it is to be expected that the stimuli used in Experiment 2 would have been longer than those which Wheeldon presented to her subjects. Both the functor/contentive distinction and the number of syllables in a word are implicated in the likelihood of its being recognised before acoustic offset (Bard et al.1988; Grosjean 1980).

The dependent variable in the analysis just described involved a perfect match between target and response in the sense that if, for example, a subject hearing the target word *photographs* offered the response *photograph*, s/he would not be credited with a correct response. Stanners et al.(1979) have reported results which suggest that words containing regular inflectional suffixes do not have independent lexical representations but are recognised via a process of morphological decomposition which leads to the accessing of the word stem. Thus, in the example above, a subject would recognise the word *photographs* by stripping the inflectional morpheme *-s* and accessing the stem *photograph*. Although Stanners et al.'s study was conducted in the visual modality, it is not unreasonable to suggest that a similar storage mechanism would be involved in auditory word recognition, particularly since some inflectional suffixes, in view of their word-final position, are especially prone to deletion in connected speech (Brown 1977). Accordingly, all inflectional suffixes in both the targets and the responses were disregarded, and the analysis was re-run. Cell means and ANOVA results from this analysis are presented in Appendix H. The results did not differ substantially from those just described; cell means were generally one or two percentage points higher than those presented above, but once again Version and Stress were the only significant main effects, with a significant Version by Speaker interaction.

The percentage of stimuli which subjects failed to recognise by acoustic offset is high in this experiment -- 37% overall, with 50% unidentified in the Spontaneous condition, and 24% in the Read condition. The discrepancy between the Read and Spontaneous means casts doubt on the assumption, implicit in most psycholinguistic experiments, that the read speech materials commonly employed are representative of the style of speech to which hearers are usually exposed.

However, it is interesting to note that the level of intelligibility in the Read condition of this experiment differs markedly from that observed in the No Context condition of Experiment 1: 76% in the current experiment versus 94% in Experiment 1 (see Chapter 4, Section 3.2.2.) (3). No doubt this difference may be explained partly in terms of the selection constraints used in the two experiments: all the Experiment 1 stimuli were trisyllabic, while a large proportion of the test words in Experiment 2 were bisyllabic. Various writers have linked intelligibility to stimulus length, whether measured in milliseconds (Mehler, Segui and Carey 1978; Fowler and Housum 1987; Bard et al.1988) or number of syllables (Grosjean 1980; Bard et al.1988).

A further aspect of the explanation may perhaps be found in the manner in which the read speech materials were produced. Pedlow and Wales (1987) demonstrated that speakers are sensitive to the wider context of their utterances: when their speakers produced lists, utterances in predictive contexts and utterances in non-predictive contexts, the presence of predictive context enhanced intelligibility only if the utterances had originally been produced in blocks of the same type (i.e. all lists, all predictive contexts, all non-predictive contexts). Pedlow and Wales suggested that their speakers were sensitive to the supra-sentential structure in which the utterances were produced. They found that utterances were overall less intelligible when stimuli were blocked (i.e. when the speakers were provided with supra-sentential information). It is possible that a similar factor is responsible for the difference in intelligibility between Experiments 1 and 2. The stimuli for Experiment 1 -- 48 unrelated sentences -- were produced at a single session. By contrast, the speakers in Experiment 2 took part in the conversational recording session, then, a few days later, recorded the read materials. Although the conversational recordings were each of one hour's duration, a limited number of topics was covered in each case. Thus it is likely that speakers were able to make connections between transcribed sentences (even though these were not read in the order in which they appeared in the original conversations) and indeed to recall the original context in which they were uttered, producing effects of comparatively low intelligibility similar to those induced by Pedlow and Wales' blocking procedure.

Whatever the reason, intelligibility differences between the Read materials in this and the previous experiment were found. It was suggested in the introduction that the ease with which subjects in Experiment 1 processed unstressed syllables might have been due to the high level of intelligibility of the materials used, and that the inclusion of spontaneous materials in Experiment 2 might reveal greater processing difficulties associated with unstressed syllables: in other words, it was hypothesised that there would be a link between level of intelligibility and the speaker's reliance on stressed syllables. This prediction seems to be borne out by the fact that a stress effect (i.e. more words with initial stress identified than words with unstressed initial syllables) was found for read as well as spontaneous materials in the current experiment. Note that in Experiment 1 stress effects were found for a number of dependent variables which, it was suggested, were related to the lexical access phase of processing, but that by the time subjects were able to recognise the test words, the advantage associated with words stressed on the first syllable had disappeared. In the present experiment, lexical stress effects were apparent even using this lexical retrieval measure, and it is interesting to note, in the light of the hypotheses under consideration, that the read materials were less intelligible overall than those in Experiment 1.

2.2.2. The Phonological Access Hypothesis

The Cohort Model states that hearers use the initial portion of words to achieve access between input and lexicon, while Grosjean and Gee (1987), among others, have suggested that stressed syllables are used to make contact with the stored representations of words. As in Chapter 4, several dependent variables were used to examine the issues relating to this hypothesis. First, if the strong version of the stressed syllable model is correct, and only stressed syllables are used to access the mental lexicon, then subjects ought to demonstrate an ability to distinguish between stressed and unstressed syllables prior to their recognition of the words which contain them. Second, if 'word onset' models such as Cohort are correct, the responses offered after subjects have heard some initial portion of the stimulus (say, the portion up to and including the first vowel) should correspond relatively closely to the input, irrespective of the stress level of the initial portion. The stressed syllable model, on the other hand, predicts that while there will be a close match for stressed syllables this will not be the case for unstressed syllables.

For the first three analyses reported below, it was necessary to identify the gate at which subjects had available to them the initial (C)CV of the input (see Chapter 4, Section 2.2.1). Because this was a time-consuming exercise in Experiment 1, which contained only 48 stimuli, it was decided that only a subset of the 576 stimuli of Experiment 2 would be analysed.

For each of the six speakers in Experiment 2, sixteen word types (8 stressed on the first syllable and 8 stressed on the second) were selected; since each type was represented by two tokens (one Read and one Spontaneous), for each speaker, 32 tokens were included in the experimental materials, resulting in 192 stimuli in all. Across the two stress categories, words were matched for number of syllables. The words which were selected included those which had proved most difficult to recognise in the analyses reported in Section 2.2. of this chapter, but as far as possible the subset was selected so as to be representative of the complete sample used in Experiment 2 in terms of intelligibility. The characteristics of the subset with respect to this and other variables are discussed below.

Word frequencies (Kucera and Francis 1967) for this subset were input to a 2-way (Speaker by Stress) ANOVA. Neither of the main effects nor the interaction reached significance, in parallel with the result of the analysis of the complete set of stimuli. Cell means are shown in Table 5.7 and ANOVA results in Table 5.8. A further variable, position in sentence, which was identical across levels of the Version variable, did not differ significantly across Stress categories, but inter-speaker differences were found ($F(5,84) = 3.70$, $p = 0.0045$).

Tables 5.9 and 5.10 present cell means for the lengths of the word tokens and the results of an ANOVA comparing them. Spontaneous words were shorter than read words. Words stressed on the second syllable were longer than those stressed on the first syllable. The only Speaker difference which reached significance was between speakers 4 and 6 ($p < 0.01$, by-subjects Scheffé test). Once again, the pattern of results for the subset reflects the characteristics of the complete set of words presented in Experiment 2.

The intelligibility scores of the words in the subset were analysed: the results of this analysis are shown in Tables 5.11 and 5.12 and illustrated in Figure 5.1. Once again, results for the subset show the same general trends as the full stimulus set used in Experiment 2 (see Figure 5.1 for a comparison of the means for the complete sample and the subset). One difference between the ANOVA results in Table 5.12 and the corresponding results for the full sample is the presence in Table 5.12 of a Stress by Version interaction which was not found in the full sample (see Table 5.6). Inspection of the means in a Scheffé test revealed that the stress difference was significant for both the spontaneous and read sub-samples, but to a lesser degree in the former than in the latter ($p<.05$ and $p<.01$, respectively).

Table 5.7: Frequencies of a subset of test words, Experiment 2

STRESSED SYLLABLE:	First	Second	Mean
SPEAKER			
1	117	130	124
2	75	93	84
3	128	109	119
4	85	89	87
5	156	117	137
6	87	53	70
Mean	108	99	

Table 5.8: ANOVA for data in Table 5.7

Source	F
Stress	(1,84) 0.14 n.s.
Speaker	(5,84) 0.75 n.s.
Speaker X Stress	(5,84) 0.16 n.s.

Table 5.9: Lengths (in milliseconds) of a subset of test words, Experiment 2

VERSION:	Read			Spontaneous					
							Stress Mean		
STRESS:	1st	2nd	(Mean)	1st	2nd	(Mean)	1st	2nd	(Mean)
SPEAKER									
1	459	504	(482)	370	426	(398)	415	465	(440)
2	356	478	(417)	286	409	(348)	321	444	(383)
3	463	489	(463)	389	384	(387)	413	437	(425)
4	453	517	(485)	451	502	(477)	452	510	(481)
5	367	460	(414)	391	373	(382)	379	417	(398)
6	335	338	(337)	321	317	(319)	328	328	(328)
Mean	401	464	(433)	368	402	(385)	385	433	(409)

Table 5.10: ANOVA for data in Table 5.9, related to Speaker, Stress and Version

Source	F
Version	(1,84) 21.16 **
Stress	(1,84) 5.90 *
Speaker	(5,84) 4.58 **
** $p < 0.01$	* $p < 0.05$

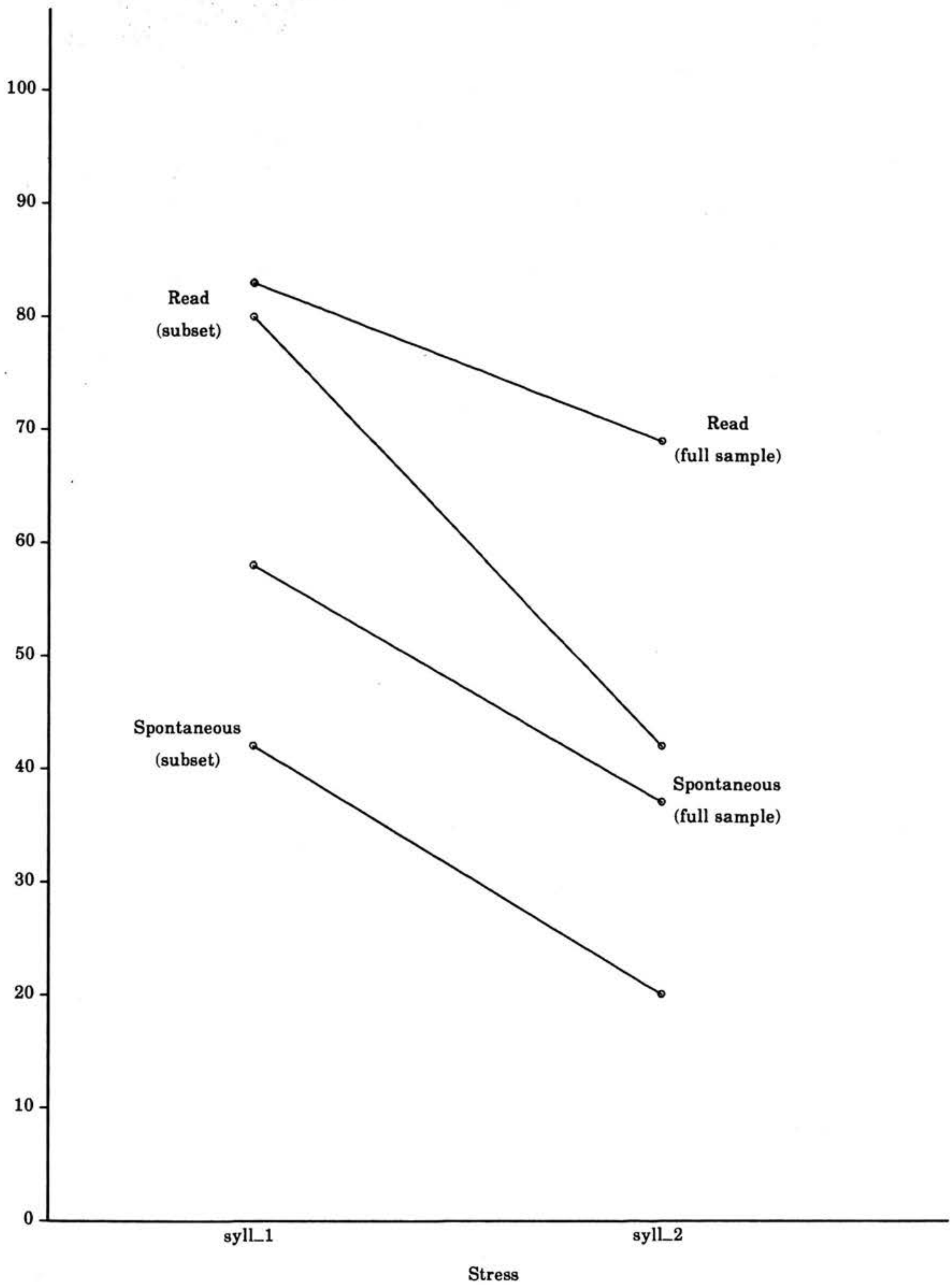
Table 5.11: Percentage of a subset of words used in Experiment 2 which were identified by acoustic offset

VERSION:	Read	Spontaneous	Mean
STRESS:			
Syll-1	80	37	59
Syll-2	42	20	31
Mean	61	29	

Table 5.12: ANOVA for data in Table 5.11, relating to Version, Stress and Speaker

Source	F
Version	(1,84) 77.66 **
Stress	(1,84) 20.06 **
Speaker	n.s.
Version X Stress	(1,84) 4.85 *
** $p < 0.01$	* $p < 0.05$

Figure 5.1:
Percentage of trials resulting in identification by offset, Experiment 2 (full sample vs subset)



Once the subset had been identified, each of the 192 stimuli in the subset was examined by the experimenter using the Audlab signal processing software (Terry et al, 1986) in order to identify, using visual and auditory information, the gate at which subjects would have been hearing most, if not all, of the initial (C)(C)V of the stimulus.

In the selection process, only the factors described above were taken into account, with the result that different subjects were unevenly represented in the subset. In view of this imbalance, by-subjects ANOVAs would be at best difficult to interpret. Therefore, in the analyses based on the subset of Experiment 2 materials, only by-materials ANOVAs were conducted. Of the by-subjects and by-materials methods of analysis, the latter is generally found to be the more conservative.

2.2.2.1. Sensitivity of subjects to the stress pattern of the input

It was argued in Chapter 4 that a prerequisite of the strong version of the stressed syllable model is that subjects should be able to distinguish stressed from unstressed syllables before recognising the words which contain them (although the absence of such an ability did not rule out a weaker version of the theory). This ability was tested in the analysis summarised in Tables 5.13 and 5.14. For the subset of 192 stimuli identified in the previous section, the responses offered at the gate at which subjects heard the initial (C)(C)V were examined. Sensitivity to stress pattern would be demonstrated if subjects offer responses consistent with the stress level of the initial (C)(C)V: initially-stressed responses when the stimulus syllable was stressed, and initially-unstressed responses when the stimulus syllable was unstressed.

Table 5.13: Percentage of responses with initial stress offered in Experiment 2 in response to the initial (C)(C)V of the stimulus.

VERSION:	Read	Spontaneous	Mean
STRESS			
Syll-1	98	98	98
Syll-2	91	93	92
Mean	95	96	

Table 5.14: ANOVA for data in Table 5.13, relating to Version and Stress

Source	F
Version	n.s.
Stress	(1,84) 3.68 **
** p < 0.01	* p < 0.05

According to this analysis, hearers do indeed display some sensitivity to the stress level of the stimulus, since they offer more responses with initial stress when the stimulus itself is stressed than when it is unstressed; but the low level of accuracy in the Syll-2 condition (subjects were wrong 92% of the time) suggests that if the stressed syllable model is operating, it is highly inefficient. However, it is possible that the results of this analysis do not adequately reflect subjects' prosodic processing abilities. In the coding scheme adopted for the analysis, monosyllabic words were considered to be initially stressed; yet there is a tendency on the part of subjects in gating experiments to offer monosyllabic responses at the early gates (Grosjean 1980). Since this behaviour may have little to do with hearers' perceptual abilities, the result observed in this analysis may be artefactual. Therefore, monosyllabic responses were excluded from the data and the analysis was re-run. The results of the re-analysis are shown in Tables 5.15 and 5.16.

Once again, this analysis certainly demonstrates some awareness of stress level on the part of hearers, since the stress variable reached significance. However, subjects made incorrect judgments about stress level in 31% of cases. If the stressed syllable model were correct, hearers would be erroneously initiating lexical access once in every three syllables. While this would seem to be a rather inefficient strategy, the result appears to indicate that hearers are making use of some information associated with stressed syllables, and account for this behaviour in models of word recognition. However, CM1 embodies no mechanism for handling prosodic information of this kind.

Table 5.15: Number of polysyllabic responses with initial stress, Experiment 2 (bracketed figures indicate percentages).

VERSION:		Read	Spontaneous	Total
Stress	Syll-1	66 (93%)	56 (91%)	122 (92%)
	Syll-2	33 (65%)	25 (64%)	58 (89%)
	Total	99 (81%)	81 (81%)	

Table 5.16: ANOVA for data in Table 5.15, relating to

Source	F
Version	n.s.
Stress	(1,218) 30.18 **
** $p < 0.01$	* $p < 0.05$

2.2.2.2. Phonological match between initial portion of input and responses

By contrast with the stressed syllable model, other theories emphasise the importance of the initial portions of words in the process of access between input and lexicon. In this section, the degree of match between the initial portions of the stimulus and the responses offered by subjects is considered.

Models such as Cohort, which claim that the initial portion of the word is used to access the set of candidates in the mental lexicon, make no explicit distinction between stressed and unstressed syllables. Thus, the responses offered by subjects hearing the initial (C)CV should match the stimulus equally well, irrespective of the stress level of the syllable. The stressed syllable model, on the other hand, predicts that the match should be close for stressed stimuli and poor for unstressed syllables.

The analyses reported in Tables 5.17 and 5.18 and illustrated in Figure 5.2 refer to the number of responses which matched the syllabic stimulus (the initial (C)(C)V of the test word) for that initial portion. For this analysis, cell means for individual speakers are presented because they are relevant to the discussion.

Table 5.17: Percentage of responses in Experiment 2 which matched the stimulus for initial (C)(C)V (data from the gate at which this portion was first available to subjects).

VERSION	Read			Spontaneous		
STRESS	Syll-1	Syll2	(Mean)	Syll-1	Syll-2	(Mean)
SPEAKER						
1	16	25	(21)	19	16	(18)
2	56	13	(34)	31	16	(24)
3	44	22	(33)	3	13	(8)
4	47	16	(32)	19	6	(13)
5	75	25	(50)	21	16	(19)
6	34	16	(25)	31	6	(19)
Mean	45	20	(33)	21	12	(16)

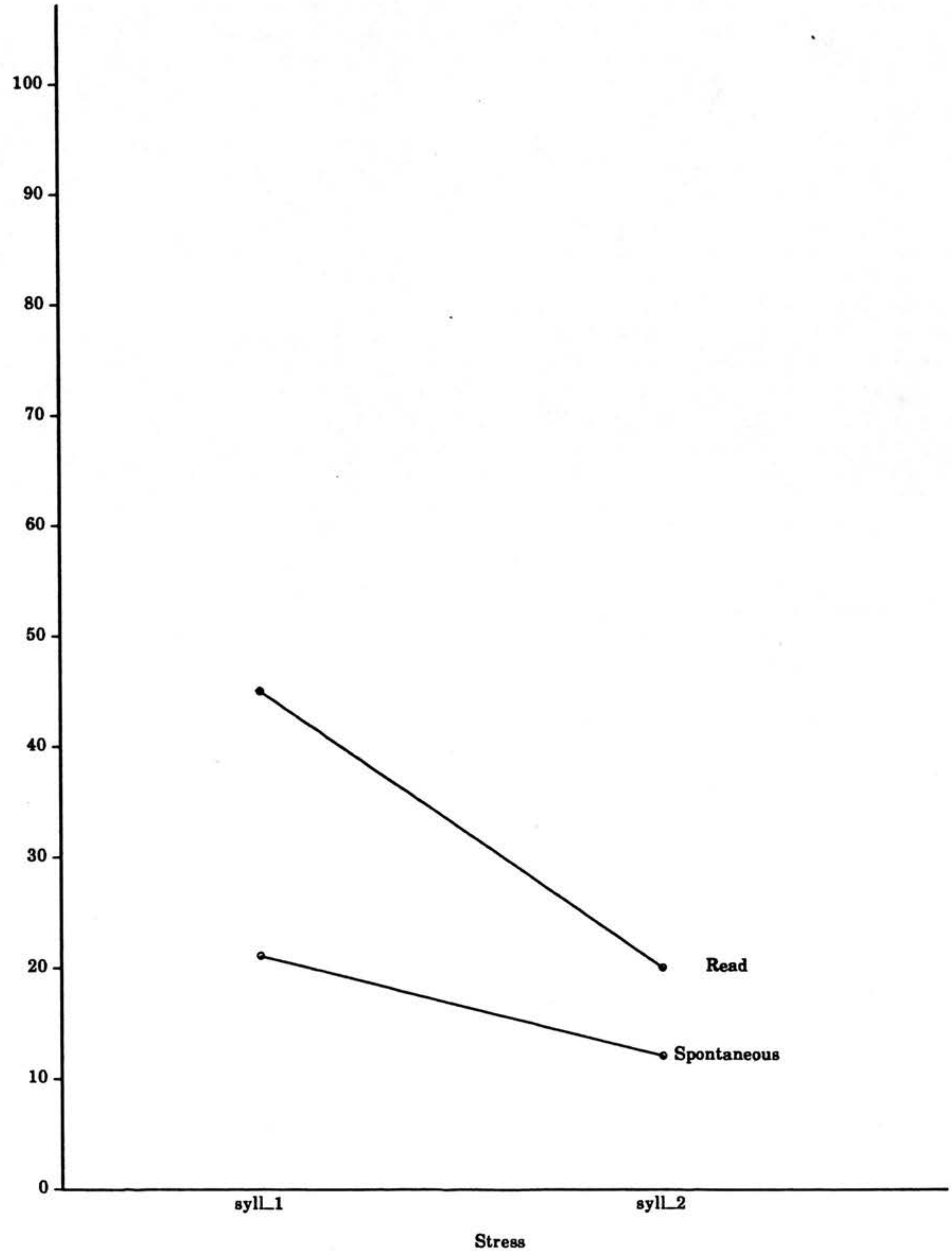
Table 5.18: ANOVA for data in Table 5.17, relating to

Source	F
Version	(1,84) 19.19 **
Stress	(1,84) 14.46 **
Version X Stress	(1,84) 5.62 *
** $p < 0.01$	* $p < 0.05$

All of the means were rather low, reflecting a poor ability on the part of hearers to analyse the initial portions of words. Although CM1 predicts that on hearing the initial syllable of a word hearers should be able to respond with items drawn mainly from the same word-initial cohort as the stimulus, these subjects fail to do so in a substantial number of cases.

The Version and Stress effects reached significance, but these variables interacted. Post-hoc comparisons of means in a Scheffé test revealed that only the difference between the two Read means and between Read/Syll-1 and Spontaneous/Syll-2 were significant ($p < 0.05$)). These results are difficult to explain in terms of CM1, which does not allow for initial syllables of differing levels of difficulty in terms of bottom-up processing.

Figure 5.2:
Percentage of responses consistent with initial (C)(C)V of stimulus, Experiment 2 (Stress X Version interaction)



However, this interaction requires further examination. In this analysis, the stress effect is significant in the read sample; the same trend is found for the spontaneous sample, but the result is not significant. Examination of the spontaneous cell means for all six speakers reveals that only one speaker's results run against the general trend. For speaker 3, only 3% of trials in which a stressed word-initial syllable was heard elicited accurate responses. When a further ANOVA was run excluding Speaker 3's data, both Version and Stress reached significance, but did not interact. Thus, there is a strong trend in these results for words stressed on the first syllable to elicit more accurate responses than words stressed on the second syllable, but Speaker 3's stimuli do not adhere to this trend.

2.2.2.3. Cohort entry

The analysis presented in Section 2.3.2. failed to support the claims of CM1 in that hearers frequently failed to enter the word-initial cohort after hearing the initial syllable of a word. However, it may be that this analysis was based on too simple-minded an interpretation of the predictions of CM1. For example, perhaps hearers need to perceive any syllable-final consonants before they can process the syllable. To circumvent this potential problem, the next analysis does not sample the data at a predetermined point in the word, but instead examines the data as a whole in order to discover the point at which subjects did in fact enter the word-initial cohort. The results of this analysis are presented in Tables 5.19 to 5.22. For the purposes of this analysis the cohort entry point is defined (as in Chapter 4) as the point at which a subject begins to offer responses consistent with the initial (C)(C)V of the input and does not subsequently offer responses from a different cohort.

The first dependent variable examined was the percentage of trials on which the appropriate cohort was entered by the acoustic offset of the word (see Tables 5.19 and 5.20). Subjects were more likely to fail to enter the appropriate cohort prior to the word's acoustic offset if they were hearing a spontaneous, rather than a read, token.

A similar disadvantage was found for words stressed on the second syllable compared with initially stressed tokens.

When the results of this analysis are compared with those reported in the previous section, it becomes clear that there is a certain delay between the availability of a particular piece of acoustic-phonetic information and subjects' ability to process it. In many cases, hearers begin to offer responses consistent with the initial CCV only when additional right context is available. The analysis reported in section 2.3.2. revealed that in the majority of cases subjects failed to enter the appropriate cohort when they heard the acoustic material for the initial (C)(C)V. However, the present analysis shows that by the acoustic offset of a stimulus, the appropriate cohort has been entered in the majority of cases.

Table 5.19: mean percentage of words for which the appropriate cohort was entered by acoustic offset (Experiment 2; subset)

VERSION	Read	Spontaneous	Mean
STRESS:			
Syll-1	85	61	73
Syll-2	77	40	58
Mean	81	51	

Table 5.20: ANOVA for data in Table 5.19, relating to Version, Stress and Speaker

Source	F1
Version	(1,94) 52.59 **
Stress	(1,94) 7.22 **
** p < 0.01	* p < 0.05

The second analysis relates to the amount of the word which subjects needed to hear before they were able to enter the appropriate cohort. As Tables 5.21 and 5.22 show, hearers can enter cohorts more quickly for Read words than for Spontaneous words and for words stressed on the first syllable than for words stressed on the second syllable. Note that the magnitude of the difference is probably underestimated by the analysis, which uses number of gates heard as the dependent variable. In fact, a given number of gates from a Read word must generally represent fewer segments of the word than the same number of gates from a spontaneous token, given the length difference between levels of the Version variable reported above. A similar argument applies to the Stress variable, as noted in Chapter 4. No interactions were significant in this analysis.

Table 5.21: Mean number of 50 msec gates heard before cohort entry (Experiment 2; subset)

VERSION	Read	Spontaneous	Mean
STRESS:			
Syll-1	3.8	4.9	4.4
Syll-2	5.4	6.2	5.8
Mean	4.6	5.6	

Table 5.22: ANOVA for data in Table 5.21, relating to Version, Stress and Speaker

Source	F1
Version	(1,84) 37.00 **
Stress	(1,84) 5.34 *
Speaker	n.s.
** $p < 0.01$	* $p < 0.05$

2.2.2.4. Summary of evidence relating to the Phonological Access Hypothesis

The analyses presented in this section have related to the question of the unit which hearers use to make contact with the mental lexicon. Two possibilities were considered: the initial syllable and the stressed syllable. The high proportion of cases in which subjects failed to distinguish stressed from unstressed syllables does not support the strong version of the hypothesis that stressed syllables fulfill this role; however, the result is not inconsistent with a weaker formulation of the theory, and hearers did exhibit some awareness of prosodic structure in their responses. On the other hand, the predictions of the cohort model, which claims that initial syllables are used to access lexical representations, receive little support: subjects continue to offer responses from inappropriate cohorts for some time after hearing the initial syllable of the stimulus; and, possibly more damaging to the theory, different tokens of the same words elicit dramatically different cohort entry points.

2.2.3. Summary and Discussion of Experiment 2

The results of Experiment 2 are summarised in Table 5.23.

Table 5.23: Summary of significant results of Experiment 2

Notes:
1. For the first analysis reported, results were significant for F1, F2 and MinF′; elsewhere, only the F2 analysis was conducted (see text).
2. The Speaker variable has been omitted.

EFFECT:	Version	Stress	Interaction
Early recognition hypothesis			
Percentage of trials in which testword was isolated by offset (full data set)	*	*	
Percentage of trials in which testword was isolated by offset (subset)	*	*	*
Phonological access hypothesis			
Percentage of responses with initial stress		*	
Number of polysyllabic responses with initial stress		*	
Percentage of cases in which initial CCV of stimulus and response matched	*	*	*
Percentage of cohort entries before offset	*	*	
Number of 50 msec gates before cohort entry	*	*	

In Experiment 2, three independent variables were manipulated: Version, Stress and Speaker. Of these, the Speaker variable produced the least interesting results. As we might expect, there is some evidence of a difference in intelligibility among the six speakers. But, despite occasional interactions between Speaker and the other two experimental variables, the same general trends were observed for most speakers, suggesting that the results are generalisable rather than the result of some idiosyncratic characteristic of certain speakers.

Part of the reason for conducting Experiment 2 with a large number of word tokens was to assess the intelligibility of the corpus with a view to selecting a subset of materials for further analysis and experimentation. Some interesting results emerged from the intelligibility analysis. First, the difference between the read and spontaneous speech samples reached significance. There is no mechanism for dealing with these differences between tokens of the same word type in CM1. The later version of the Cohort Model allows for such a possibility, however. The extent to which this later model explains the results will be considered in Chapter 6.

The second interesting aspect of the results of the intelligibility analysis was the significant stress effect: words with stressed initial syllables were more likely to be identified than words with unstressed initial syllables. Except for one speaker, this result was found for both read and spontaneous speech samples. This result is the more interesting because of the failure to find any stress effect for the same dependent variable in Experiment 1. A possible explanation is that the overall level of intelligibility, which was much higher in Experiment 1 than in the read speech sample of Experiment 2, allowed the value of stressed syllables as islands of reliability to emerge in the latter but not in the former.

The main hypothesis which was examined in Experiment 2 was the phonological access hypothesis. The aim of the experiment in this respect was to distinguish between the claims of the Cohort Model, which emphasises the importance of word onsets in lexical access, and those of a stressed syllable model such as that proposed by Grosjean and Gee (1987), which assigns a central role to stressed syllables. Although subjects could only identify the stress level of a syllable in about a third of cases, the accuracy of their responses when they heard the initial syllable of a stimulus was influenced by the stress level of the item. By contrast, the predictions of CM1 with respect to phonological access received little support in Experiment 2. First, presenting subjects with the initial syllable of stimuli did not enable them, as the model predicted it should, to constrain their responses to only those words which were members of the same word-initial cohort as the stimulus. Second, it was clearly the case that some word-initial syllables were easier to process than others: subjects had comparatively little difficulty in entering the word-initial cohort when they were hearing stressed initial syllables from read tokens, but their performance was poor on spontaneous tokens with unstressed initial syllables. CM1 cannot account for this finding because it has no mechanism for dealing with the variability of word tokens.

It is interesting to note, in connection with this finding, that word recognition experiments which have been cited in favour of the cohort model have tended to use read speech materials, and that these materials have either been monosyllabic content words, which would tend to be stressed, or else polysyllabic items in which stress was not explicitly controlled. For example, the materials used by Marslen-Wilson and Tyler (1980) consisted of 81 content words, 77 of which were monosyllabic, while Marslen-Wilson and Welsh (1978) used polysyllabic words and noted (1978:41) that in their 'first-syllable' condition, which they claimed to demonstrate the importance of initial syllables, the majority of the words bore initial stress. In view of this fact, and of the results reported here, it is hardly surprising that experimental results have tended to support the view of processing contained in CM1.

The conclusions which can be drawn from Experiment 2 are limited by the fact that the words were excised from context. While it was necessary to present the stimuli in this manner in order to isolate the effects of bottom-up processing from those of contextual factors, the hypotheses outlined above cannot be evaluated in full until the analyses relating to words heard in context have been presented. First, the early recognition hypothesis refers to polysyllabic words and to words heard in context. Since the tokens presented in Experiment 2 were originally produced in context, and since it is well known that speakers alter intelligibility of word tokens in response to contextual factors (Lieberman 1963; Wheeldon 1985; Pedlow and Wales 1988) it may be important for the words to be heard in their original contexts in order for them to be perceived correctly. Second, the effects of stress on lexical access can be better evaluated on the basis of data from words heard in context because hearers use the prosodic information in preceding sentence fragments to predict the location of upcoming stresses (Shields, McHugh and Martin 1973; Cutler 1976); perhaps the task in Experiment 2, therefore, was somewhat artificial, a factor which may account for subjects' poor ability at distinguishing stressed from unstressed syllables.

3. Experiment 3

Experiment 2 provided a measure of subjects' bottom-up processing of read and spontaneous word tokens of varying stress patterns. Experiment 3, in which the same independent variables are manipulated for a subset of the materials in Experiment 2, seeks to examine the early recognition and phonological access hypotheses on the basis of responses produced when contextual information was available to the hearer.

3.1. Method

3.1.1. Materials

The stimuli used in Experiment 3 consisted of the subset of the 192 items from Experiment 2 which were analysed in Sections 2.3.1. to 2.3.3. The principles governing their selection are discussed in Section 2.3., and the test words themselves are listed, with sentential contexts, in Appendix I.

Test words were all presented following an ungated version of their left sentence contexts and were gated, in the manner described in Experiment 2, in 50 millisecond increments. In addition, up to three words following the test word were gated, the precise number being constrained by the length of the sentence; for example, since the testword was penultimate in some sentences, only one word following it was gated.

3.1.2. Subjects

Sixteen undergraduates at the University of Edinburgh participated as part of a course requirement.

3.1.3. Design

Each subject took part in two sessions lasting approximately forty minutes each. In a single session, subjects heard four tokens from each Speaker, one from each Version by Stress category, but no subject heard two tokens of the same word type. The design was Version (2 levels: Read, Spontaneous) X Speaker (6 levels) X Stress (2 levels: First Syllable Stress, Second Syllable Stress). Subjects were nested in groups defined as above, and crossed with all other variables. Stimuli were blocked according to Speaker.

3.1.4. Procedure

Subjects were instructed to listen to the stimulus and to identify the first of the gated words. Their responses were recorded as in Experiment 2. Experimental trials were preceded by four practice trials, one from each Version X Stress category. An orientation sentence preceded each Speaker-defined block. Inter-trial and inter-block intervals were as described for Experiment 2.

3.2. Results

3.2.1. The Early Recognition Hypothesis

Word onset models such as Cohort require that the majority of words be recognised by their acoustic offsets in order that the locations of the onsets of subsequent words can be identified. Indeed, Marslen-Wilson (1987) has claimed explicitly that the majority of words heard in context will be recognised by their acoustic offset. This early recognition hypothesis will be examined in this section.

The analyses presented here relate to the isolation point dependent variable. The isolation point of a word was defined by Grosjean (1980) as the point at which subjects began to offer the target word as a response, and continued to do so on all subsequent trials. The first analysis concerns the percentage of experimental trials which resulted in successful isolations. Cell means, collapsed across Speaker, for the percentage of words isolated by their acoustic offset, are presented in Table 5.24, and the corresponding ANOVA results on Table 5.25.

Table 5.24: Percentage of words isolated by acoustic offset of testword, Experiment 3

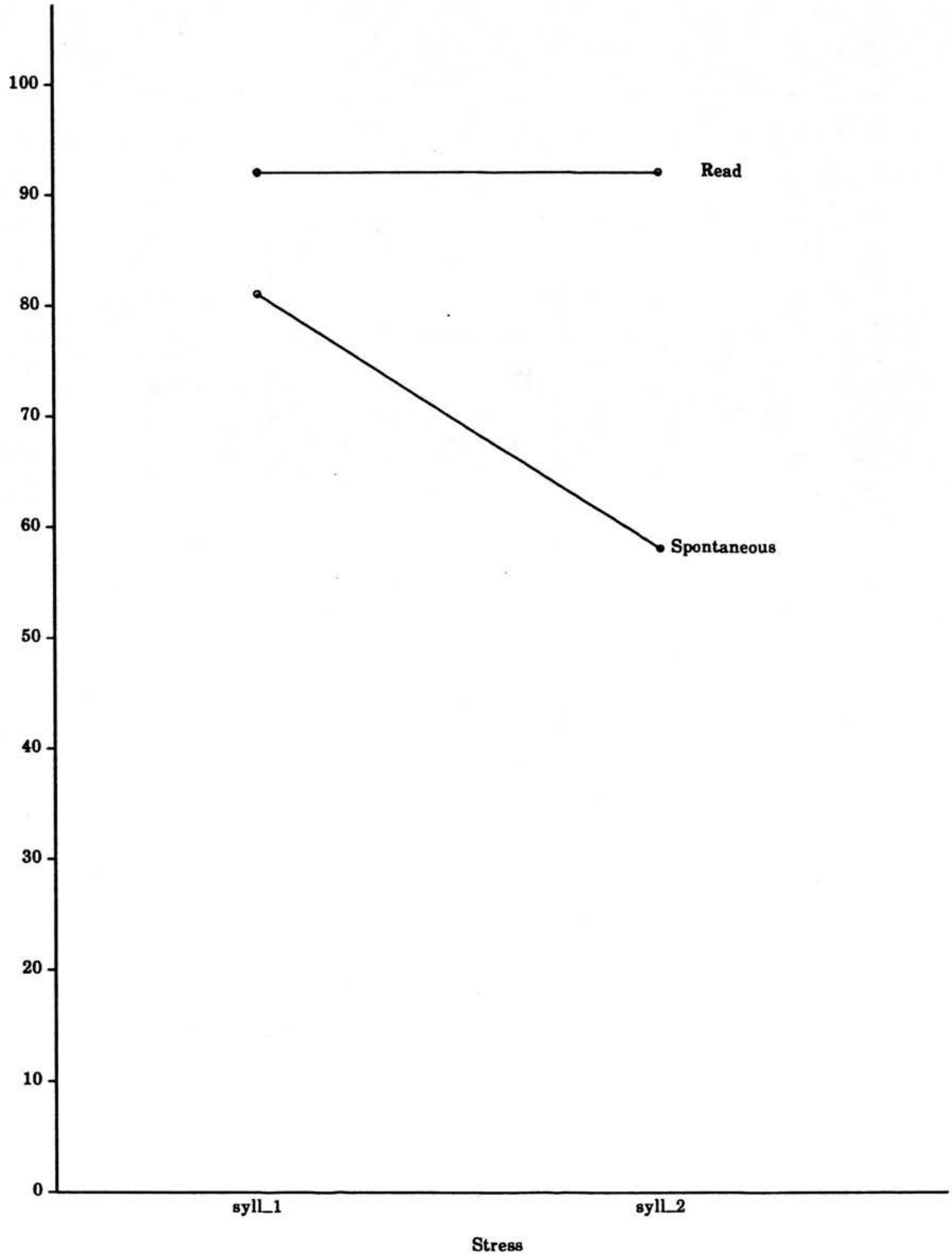
VERSION	Read	Spontaneous	Mean
STRESS:			
Syll1	92	81	86
Syll2	92	58	75
Mean	92	70	81

Table 5.25: ANOVA for results in Table 5.24

Source	F1	F2	Min F'
Version	(1,15)32.44**	(1,84)56.52**	(1,35)20.61**
Stress	(1,15)20.86**	(1,84) 6.40**	(1,93) 4.90*
Speaker	(5,75) 5.22**	n.s.	----
Version X Stress	(1,15)17.62**	(1,84)14.79**	(1,57) 8.04**
Version X Speaker	(5,75) 5.12**	n.s.	----
**p<0.01	*p<0.05		

The overall result of 19% failures to isolate words before their acoustic offsets is compatible with the findings of Bard et al (1988); but while it might be possible to argue that the words used in their experiment may have failed to be identified because they remained ambiguous, or phonetically homophonous, beyond their acoustic offset, it is highly unlikely that this argument could explain the substantial proportion of words which remained unrecognised in the current experiment, because the words in question were polysyllabic.

Figure 5.3:
Percentage of testwords identified by acoustic offset,
Experiment 3 (Stress X Version interaction)



More Read words were isolated than Spontaneous words, and more words stressed on the first syllable than words stressed on the second syllable. The Version by Stress interaction reached significance, with the spontaneous/second syllable stress mean lower than the means for other conditions ($p < 0.01$, by-subjects Scheffé test); no other comparison reached significance. Note that while the stress difference reached significance within the spontaneous sample only in Experiment 3, in Experiment 2, when the same word tokens were heard in isolation, both the read and the spontaneous sample showed a significant stress effect. The Version by Stress interaction is illustrated in Figure 5.3.

The distribution of successful recognition outcomes across the gated words is presented in Table 5.26. The N for each of the 4 conditions represented is 192 (6 speakers X 8 word types X 4 subjects).

Table 5.26: Distribution of successful isolations (Experiment 3)

	During Word 1	During Word 2	During Word 3	During Word 4	Failure to recognise	Total
Read/Syll1	176	10	2	1	3	192
Read/Syll2	176	14	2	0	0	192
Spont/Syll1	155	15	5	0	17	192
Spont/Syll2	113	22	18	4	35	192
Total	620	61	27	5	55	768

For all conditions except Spontaneous/Syll-2, the majority of words which were unrecognised by acoustic offset were identified during the following word. However, in the latter condition, subjects often needed to hear at least part of a further word before they could recognise the test word. Another notable characteristic of the Spontaneous/Syll-2 condition was the large number of complete failures to recognise testwords: 35 recognition outcomes fell into this class. The Spontaneous/Syll-1 condition shares this characteristic, although the extent of the problem is not so widespread.

Since the gating technique was used in this experiment, it is possible to provide more detailed information than has hitherto been presented in this study about the lexical retrieval process. That is, the discussion up to this point has been confined to the percentage of words which were identified by acoustic offset, a measure which could have been derived from a simpler experimental technique in which excised words were presented ungated. The advantage of using within-word gating is that it provides a measure of the amount of acoustic stimulus which it is necessary for the subject to hear in order to identify it. Tables 5.27 and 5.28 present isolation point means and ANOVA results relating to them.

The only main effect to yield a significant value for MinF′ was Version: read tokens were recognised on the basis of less acoustic input than their spontaneous counterparts. However, as in the previous analysis, this variable interacted with Stress such that although the Read means did not differ, subjects identified Spontaneous tokens more quickly if they were stressed on the first syllable than if they were stressed on the second syllable ($p < .01$, by-subjects Scheffé test). This interaction, illustrated in Figure 5.4, is consistent with the suggestion discussed in the introduction to this chapter that the stress effect was more likely to observed in less intelligible speech styles.

An interesting corollary to the results just presented emerges when the isolation point analysis is confined to those words which were isolated before their acoustic offset. The relevant data and results are presented in Tables 5.29 and 5.30: although the Version effect persists, with read tokens identified more quickly than spontaneous tokens, neither the Stress effect nor the Version by Stress interaction is observed. In other words, the stress effect is confined to those tokens which are less intelligible in the sense they remain unrecognised by acoustic offset. This finding is consistent with the hypothesis that the value of stressed syllables as islands of reliability increases in conditions of poor intelligibility.

Table 5.27: Percentage of test words heard before isolation (Experiment 3)

VERSION	Read	Spontaneous	Mean
STRESS:			
Syll-1	57	85	71
Syll-2	55	110	83
Mean	56	98	77

Table 5.28: ANOVA for data in Table 5.27, relating to

Source	F1	F2	MinF′
Version	(1,15) 44.53 **	(1,84) 67.29 **	(1,38) 26.80 **
Stress	(1,15) 10.98 **	n.s.	---
Speaker	(5,75) 7.62 **	n.s.	---
Version X Stress	(1,15) 9.93 **	(1,84) 7.50 **	(1,62) 4.27 *
Stress X Speaker	(5,75) 4.39 **	n.s.	---
Version X Stress X Speaker	(5,75) 2.92 *	n.s.	---
** $p < 0.01$	* $p < 0.05$		

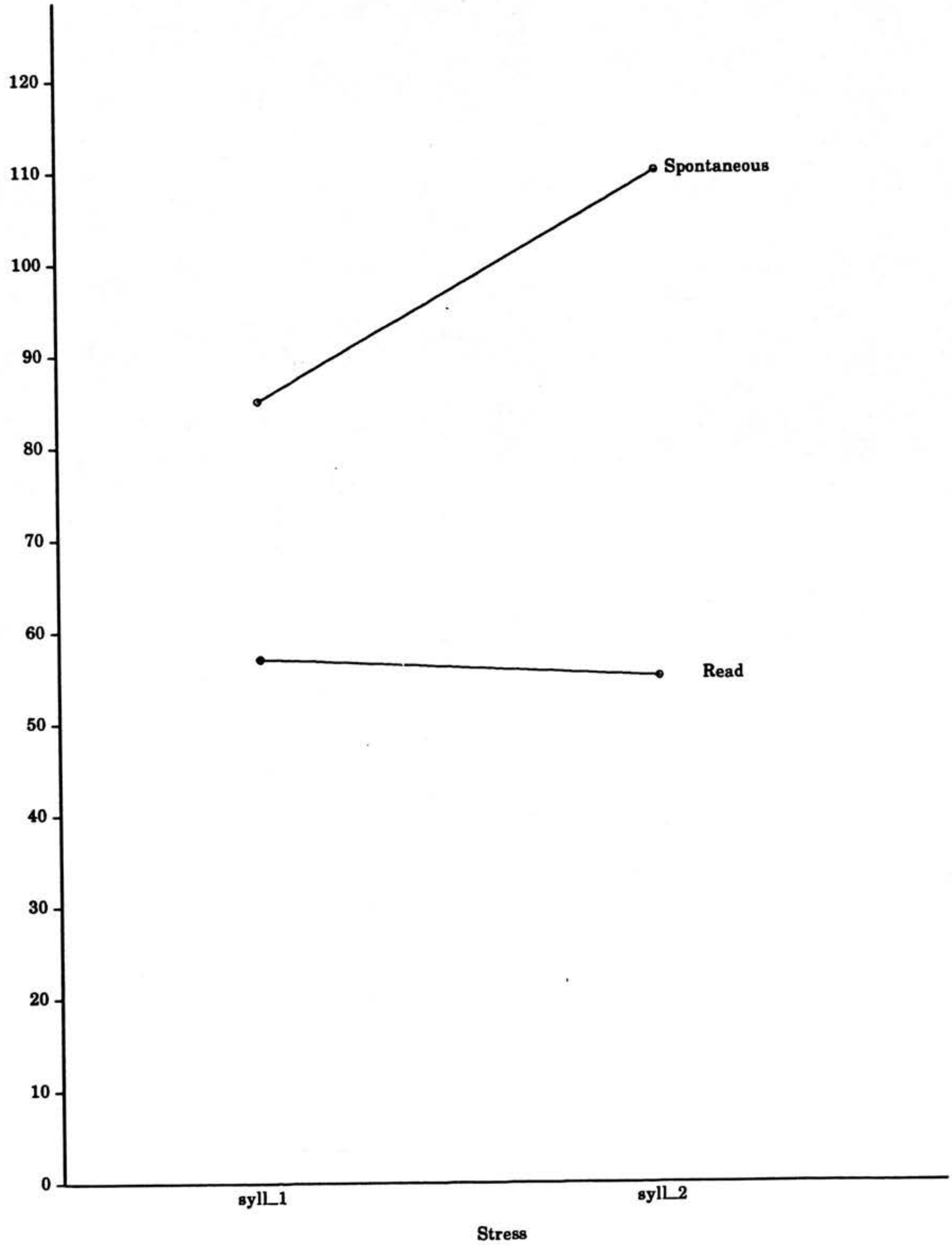
Table 5.29: Percentage of testword heard before isolation, for those words isolated before acoustic offset (Experiment 3)

VERSION	Read	Spontaneous	Mean
STRESS:			
Syll-1	47	58	53
Syll-2	48	59	54
Mean	48	59	

Table 5.30: ANOVA for data in Table 5.29, relating to Version, Stress and Speaker

Source	F
Version	(1,614) 25.41 **
Stress	n.s.
Speaker	n.s.
** $p < 0.01$	* $p < 0.05$

Figure 5.4:
Percentage of word heard before isolation,
Experiment 3 (Stress X Version interaction)



3.2.2. Phonological Access

In this section, the issue of the unit which is used for phonological access is further considered. The Cohort Model suggests that the initial portion of the word is used, while the model proposed by Grosjean and Gee (1987) claims that stressed syllables fulfill this role.

3.2.2.1. Sensitivity of subjects to the stress pattern of the input

Analysis of this same set of test words in Experiment 2 revealed some awareness on the part of the subjects of the stress level of the material they were hearing, but their relatively low level of accuracy suggested that the strategy proposed by Grosjean and Gee (1987) and others, of initiating lexical access at stressed syllables, would be an inefficient one. However, perhaps the absence of the original context in which the words were produced had a detrimental effect on subjects' ability to detect the cues for stress. Therefore, the responses offered by subjects on hearing the initial (C)(C)V of the test words were again examined, to determine whether they were able to meet the prerequisite of the model when the rhythmic context of the utterance was available to them. Results of this analysis are presented in Tables 5.31 and 5.32.

Table 5.31: Percentage of responses with initial stress which were offered at the gate at which subjects first heard the initial (C)(C)V of the stimulus (Experiment 3)

VERSION	Read	Spontaneous	Mean
STRESS:			
Syll-1	82	80	81
Syll-2	73	59	66
Mean	78	70	

Table 5.32: ANOVA for data in Table 5.31, relating to Version, Stress and Speaker

Source	F1	F2	MinF´
Version	n.s.	n.s.	---
Stress	(1,15) 36.53 **	(1,84) 21.31 **	(1,72) 13.46 **
Speaker	n.s.	n.s.	---
Version X Stress	(1,15) 6.22 *	(1,84) 5.42 *	n.s.
** p < 0.01	* p < 0.05		

Subjects' performance on Read and Spontaneous tokens did not differ significantly. There was a marked difference between the Stress means, however, with subjects more likely to offer an initially stressed response if the stimulus was initially stressed than if it was initially unstressed. There was an overall improvement in subjects' level of accuracy, compared with the corresponding analysis of the Experiment 2 data, suggesting that subjects were able to use the prosodic information in the preceding sentence fragment to predict the locations of stressed syllables. This result is compatible with the findings of other writers (e.g. Shields et al.1973; Cutler 1976). Although this result suggests that subjects in Experiment 3 were displaying an awareness of the stress pattern, an ability which is a prerequisite of the strong version of the stressed syllable model, the level of accuracy with which they performed this task was still rather poor. In the Syll-1 condition, they offered inappropriately stressed responses in approximately 20% of cases, while in the Syll-2 condition, their responses are inappropriately stressed 66% of the time. The stress difference was significant in the spontaneous sample only ($p < 0.01$, by-subjects Scheffé test).

However, it might be objected that this analysis is biassed in two ways. First, many of the responses which subjects offered were monosyllabic; this tendency is generally observed in gating tasks (see, for example, Grosjean 1980) (4). Since monosyllabic responses were coded in the analysis as bearing initial stress, this results in a preponderance of initially stressed items in the data which may in fact be due to a response bias. This suggests that a second analysis, with monosyllabic words excluded, should be performed, as in Experiment 2. Second, since the stimuli in

Experiment 3 were preceded by context, and since (as Experiment 1 has shown) context effects are discernible even in the earliest stages of the processing of words, subjects occasionally managed to recognise the testword after hearing only the initial (C)(C)V. Since contextual factors would have been partly responsible for these recognitions, it seems appropriate to exclude such responses from the analysis in order to isolate the effects of phonological processing as far as possible. Therefore, a second ANOVA was run which was confined to polysyllabic words and which excluded responses which were successful identifications of the test word. Tables 5.33 and 5.34 contain the details of this analysis.

Table 5.33: number of polysyllabic responses in Experiment 3 with initial stress (successful identifications excluded); bracketed figures indicate percentages.

VERSION	Read	Spontaneous	Total
STRESS:			
Syll-1	33 (80%)	27 (87%)	60 (83%)
Syll-2	14 (35%)	15 (54%)	29 (43%)
Total	47 (58%)	42 (69%)	

Table 5.34: ANOVA for data in Table 5.33, relating to

Source	F
Version	n.s.
Stress	(1,136) 27.84 **
Version X Stress	n.s.
** $p < 0.01$	* $p < 0.05$

Analysis of this restricted set of responses does not reveal any improvement in subjects' ability to produce appropriately stressed responses. The significant Stress result indicates some sensitivity on the part of subjects to the stress level of syllables, but the means themselves are not consistent with the stressed syllable account of lexical access, at least in its strong form. For words stressed on the first syllable, 60 of the 72 responses analysed were stressed appropriately; but for words stressed on the second syllable, only 39 of the 68 responses examined had appropriate stress patterns. It is doubtful whether this overall level of accuracy of 68% is sufficient for the strong version of the stressed syllable model to be viable. The weaker account is, however, not inconsistent with the results.

3.2.2.2. Phonological match between initial portion of input and responses

The responses produced in Experiment 3 at the gate at which subjects first heard the initial (C)CV of the stimulus were examined to determine to what extent they matched the stimulus for this initial portion. CM-1 predicts that responses will match the stimulus irrespective of stress level, while the stressed syllable model predicts a better match when the stimulus syllable is stressed.

The Version effect was significant in all three analyses, with read tokens more likely to elicit an accurate response than spontaneous tokens. The stress effect reached significance for F1 but not for F2: stress-bearing syllables produced more accurate responses than unstressed syllables.

Table 5.35: Percentage of cases in which initial (C)(C)V of stimulus and response matched (Experiment 3)

VERSION	Read	Spontaneous	Mean
STRESS:			
Syll-1	38	25	32
Syll-2	33	20	27
Mean	36	23	

Table 5.36: ANOVA for data in Table 5.35, relating to Version, Stress and Speaker

Source	F1	F2	MinF´
Version	(1,15) 32.73 **	(1,84) 6.89 **	(1,99) 5.69 **
Stress	(1,15) 4.63 *	n.s.	---
Speaker	n.s.	n.s.	---
** $p < 0.01$	* $p < 0.05$		

3.2.2.3. Cohort Entry

Particularly in the spontaneous condition, subjects frequently failed to enter the appropriate cohort after hearing a stimulus which, according to the predictions of the cohort model, should have enabled them to do so. The analyses in this section will determine the amount of input required by subjects before they *can* enter the appropriate word-initial cohort.

In Experiments 1 and 2, the cohort entry point was defined as that point at which subjects began to offer responses agreeing with the stimulus in initial (C)CV, and did not subsequently alter this aspect of their responses. The first analysis shows the percentage of experimental trials which resulted in entry of the appropriate cohort by the acoustic offset of the word.

Table 5.37: percentage of experimental trials which resulted in cohort entry by acoustic offset, Experiment 3

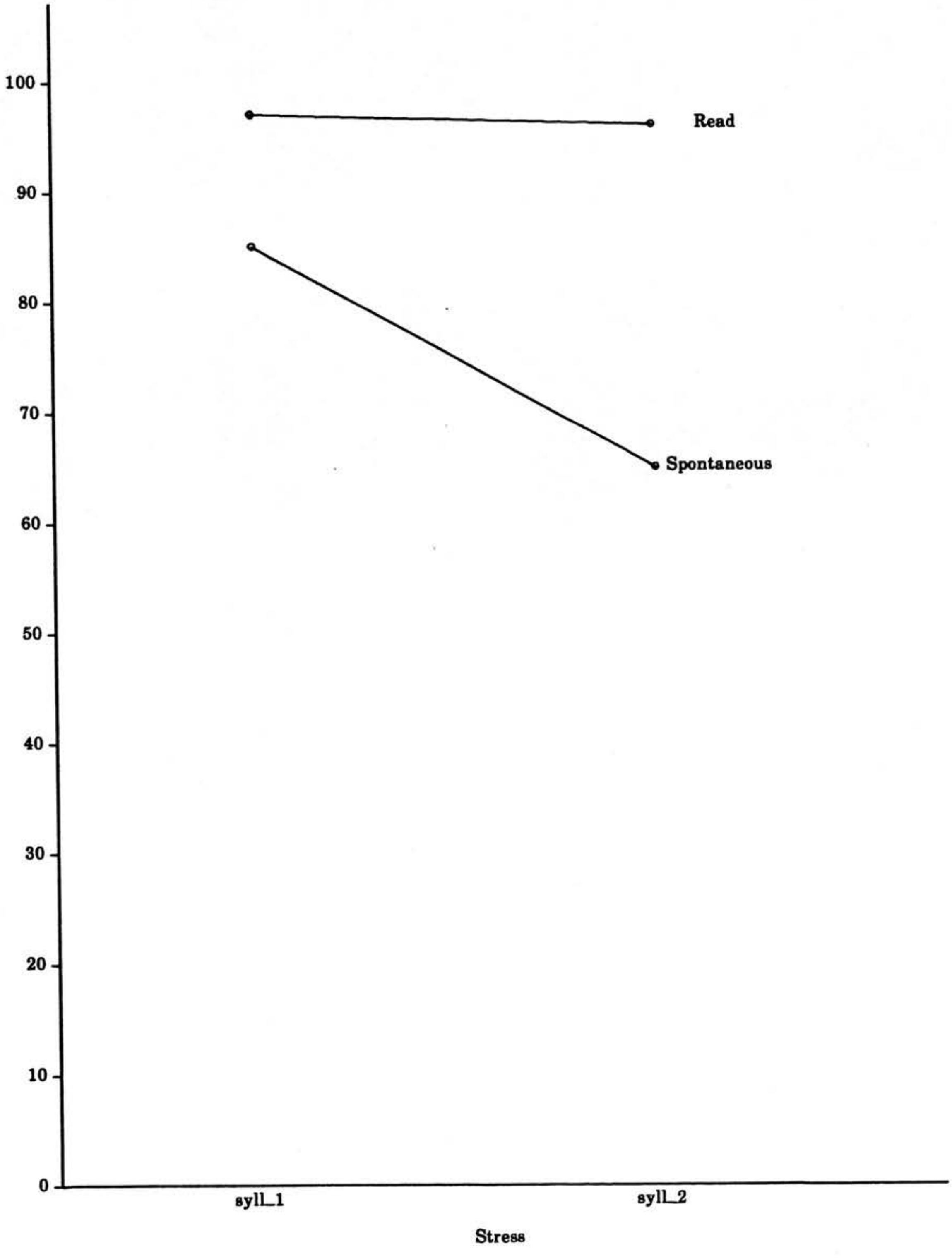
VERSION	Read	Spontaneous	Mean
STRESS:			
Syll-1	97	85	91
Syll-2	96	65	81
Mean	97	75	

Table 5.38: ANOVA for data in Table 5.37, relating to

Source	F1	F2	MinF′
Version	(1,15) 61.84 **	(1,84) 55.88 **	(1,55) 29.35 **
Stress	(1,15) 16.22 **	(1,84) 7.95 **	(1,80) 5.34 *
Speaker	(5,75) 24.73 **	n.s.	---
Version X Stress	(1,15) 24.73 **	(1,84) 9.94 **	(1,87) 7.09 **
Version X Speaker	(5,75) 3.02 *	n.s.	---
Version X Stress X Speaker	(5,75) 2.59 *	n.s.	---
** $p < 0.01$	* $p < 0.05$		

Although Version and Stress both reached significance in this analysis (with read words resulting in more cohort entries than spontaneous words, and words stressed on the first syllable resulting in more than words stressed on the second syllable) the two variables interacted (see Figure 5.5): the only reliable differences were between the

Figure 5.5:
Percentage of trials in which correct cohort was entered by acoustic offset, Experiment 3 (Stress X Version interaction)



Spontaneous/Syll-2 mean and the rest ($p < 0.01$, by-subjects Scheffé test). In other words, the stress effect was once again confined to the spontaneous condition: the stress difference in read speech, which was apparent when words were heard in isolation, disappeared.

Cohort entry points themselves were next considered. Cohort entry points for Experiment 3 are presented in Table 5.39, and the ANOVA relating to this data in Table 5.40.

Subjects needed to hear a greater portion of spontaneous tokens than read tokens in order to be able to identify the initial syllable. Both the Stress effect and the Version X Stress interaction were significant in the F1 and F2 analyses but just failed to reach significance for MinF´. Although the stress pattern of the stimulus exerts an effect in the Spontaneous condition, it was irrelevant in the Read condition.

Table 5.39: Number of gates heard before cohort entry, Experiment 3.

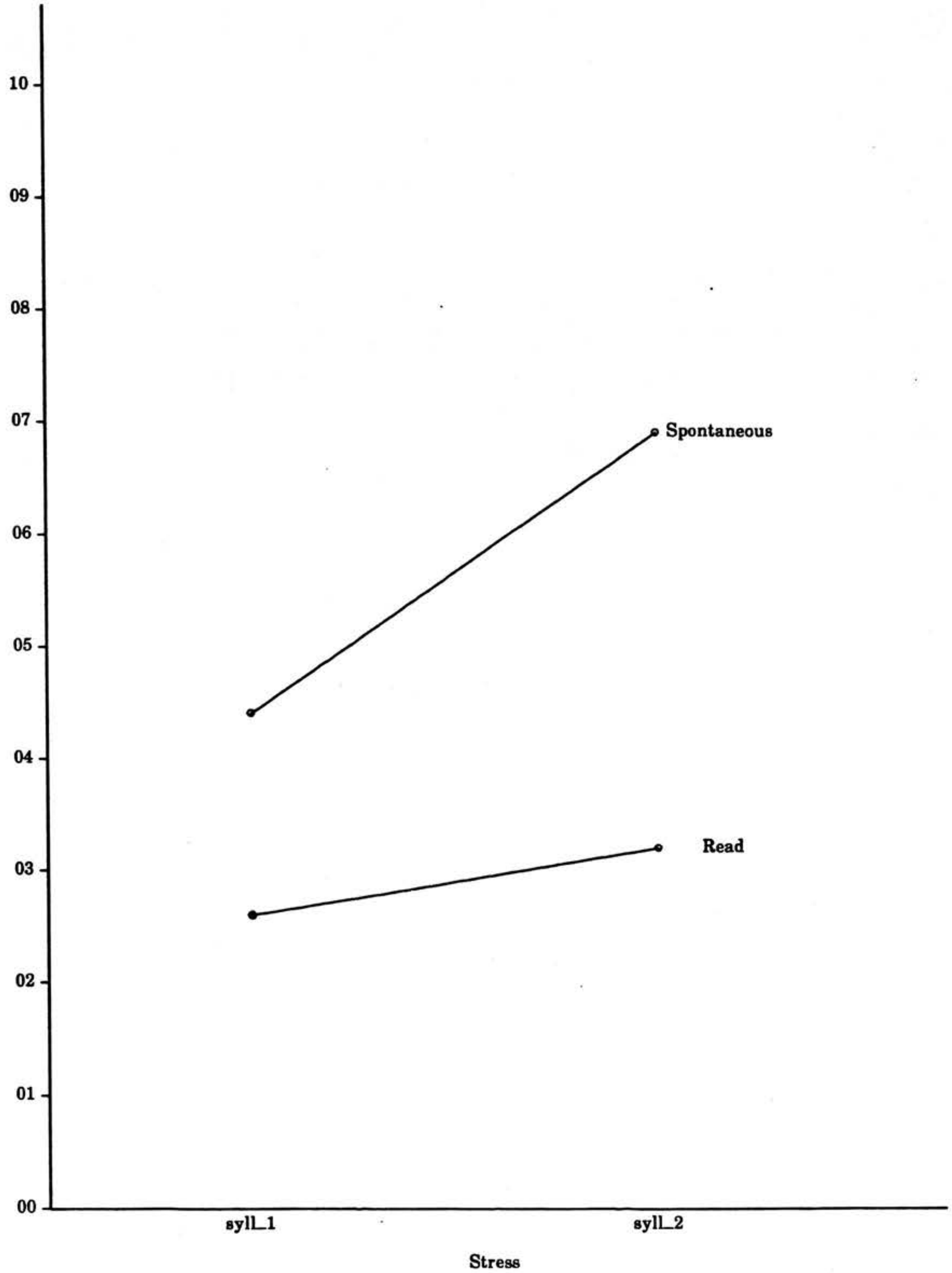
VERSION	Read	Spontaneous	Mean
STRESS:			
Syll-1	2.6	4.4	3.5
Syll-2	3.2	6.9	5.1
Mean	2.9	5.6	

Table 5.42: ANOVA for data in Table 5.41, relating to

Source	F1	F2	MinF´
Version	(1,15) 69.54 **	(1,84) 62.20 **	(1,55) 32.83 **
Stress	(1,15) 18.13 **	(1,84) 4.53 *	n.s.
Speaker	(5,75) 5.76 **	n.s.	---
Version X Stress	(1,15) 8.68 **	(1,84) 5.96 *	n.s.
Version X Speaker	(5,75) 2.44 *	n.s.	---
Version X Stress X Speaker	(5,75) 3.42 **	n.s.	---
** p < 0.01	* p < 0.05		

The only main effect to reach significance for MinF´ in this analysis was Version: once again, subjects could more readily process the initial syllables of read tokens than of spontaneous. Similarly, the stress difference once again emerged, with means for words

Figure 5.6:
Mean number of 50ms gates heard before cohort entry, Experiment 3 (Stress X Version interaction)



stressed on the first syllable lower overall than those for words stressed on the second syllable. Version and Stress interacted (see Figure 5.6), such that although the two read means were not reliably different, in the Spontaneous condition, less input was required before cohort entry when the initial syllable was stressed than when it was unstressed ($p < .01$, by-subjects Scheffé test).

3.3. Summary of Experiment 3

The results of Experiment 3 are summarised in Table 5.41.

Table 5.41: Summary of significant results of Experiment 3

Note:
* = effect significant for F1, F2, MinF′ † = only F2 analysis was run, for reasons explained in the text

EFFECTS:	Version	Stress	Interaction
Early recognition hypothesis			
Percentage of trials in which testword was isolated by offset	*	*	*
Percentage of stimulus heard before isolation	*	F1	*
Percentage of testword heard before isolation (trials on which testword isolated before offset) †	F2		
Phonological access hypothesis			
Percentage of responses with initial stress		*	F1, F2
Percentage of polysyllabic responses with initial stress †		F2	
Percentage of cases in which initial (C)(C)V of stimulus and response matched	*	F1	
Percentage of experimental trials resulting in cohort entry before word offset	*	*	*
Number of 50 msec gates before cohort entry	*	F1,F2	F1,F2

Two sets of predictions were being evaluated in Experiment 3. The first related to the lexical retrieval stage of processing. For 'word onset' models such as Cohort, the majority of words need to be recognised by acoustic offset so that the beginnings of subsequent words can be identified; this is necessary because of the absence of reliable word boundary cues in fluent speech. While the majority of Read words in Experiment 3 were recognised by their offset, this was not the case for Spontaneous words: overall, approximately one fifth of these could only be identified after subsequent words had been processed. Since the words in Experiment 3 were polysyllabic, it is unlikely that the reason for these failures was the temporary ambiguity to which Marslen-Wilson (1987) has alluded. It is interesting to note that the Version and Stress variables interacted in Experiment 3: in the Read condition, the Stress difference was irrelevant to the identification of words by their acoustic offsets, but in the Spontaneous condition, it affected both the percentage of words which were identified by acoustic offset and the amount of the word which needed to be heard before the word could be identified. These results cannot be accommodated by CM1.

In Experiment 3, both Version and Stress, the main effects under consideration, reached significance: overall, on dependent variables relating to both access and retrieval, read tokens were more efficiently processed than spontaneous, and initially stressed words were more efficiently processed than words with unstressed initial syllables. However, these two variables interacted. Only in the spontaneous sample was the Stress effect reliably present. While the results of Experiment 3 are difficult to explain in terms of the 1980 version of the Cohort Model, the stressed syllable model cannot satisfactorily account for them either, because not all of the speakers consistently exhibited the stress effect. Alternative accounts of the findings of Experiments 2 and 3 will be discussed in the next chapter.

Chapter Notes

(1) It seems quite likely that this is the case, since a large proportion of spontaneous tokens fall into the functor word classes, and these items tend to be short, in terms of

millisecond length (a variable which Bard et al. found to be highly predictive of recognition outcome). Bard et al. analysed the recognition of all the word tokens in their experimental sentences.

(2) Since a pre-test showed that speakers were embarrassed when asked to read sentences which contained gross ungrammaticalities, false starts etc, such disfluencies were not included in the transcript which speakers were asked to read.

(3) It would be useful to compare these intelligibility results with those of Grosjean (1980), who used a similar methodology. However, Grosjean reports his results not in terms of recognition outcomes, as in this study, but in terms of number of instances in which all subjects hearing a given word identify it by acoustic offset. For example, 48% of words heard in his No Context condition remained unidentified by at least one subject (note, however, that this result is collapsed for one, two and three syllable words; the materials in this study are all polysyllabic, and one would therefore expect them to be more intelligible than monosyllables). Casual inspection of the data from Experiments 1 and 2 suggests that Grosjean's materials were more similar to the former than the latter, in terms of intelligibility.

(4) Van Heuven (1988) attempted to circumvent this tendency by telling his subjects that the stimuli were all polysyllabic. Unfortunately, they frequently failed to take his advice, producing monosyllabic responses in many cases.

CHAPTER SIX
Early recognition, contextual processing and phonological access: some further evidence

Chapters 4 and 5 investigated issues arising from three hypotheses, relating to post-offset recognition, the use of context and the part of the input used to make contact with the stored representations of words. Although several questions were answered by the analyses presented in Chapters 4 and 5, other issues have been raised which were not discussed there. These issues will be examined in this chapter.

1. Early Recognition

The first hypothesis which was examined was termed the early recognition hypothesis. Models which ascribe to word onsets an important role in the lexical access process require that the majority of words heard in fluent speech should be recognised by their acoustic offset; otherwise, given the absence of word boundary cues in fluent speech, hearers would not be able to locate word onsets and would be unable to access the intended word. Indeed, Marslen-Wilson (1987:75) has explicitly claimed that all words heard in context will be recognised at or before their offsets (1). Experiment 1 suggested that this was the case for the majority of read words heard in context. In Experiment 3, polysyllabic read words heard in context were also relatively accurately identified by acoustic offset, but for a substantial number of spontaneous tokens, subjects needed to hear subsequent acoustic material before they could recognise the word. Indeed, it was often only after several subsequent words had been presented that recognition occurred. This result seems to argue against purely left-to-right models, and against those models which emphasise the importance of word onsets.

Given the failure of subjects to process words efficiently in a substantial percentage of cases, it seems appropriate to ask what factors were implicated in determining what words would be recognised 'on time'. Others who have sought to answer this question have used the statistical technique of regression. Bard et al. (1988) included in their regression analysis the following variables: length in syllables, frequency of occurrence, form class (functor vs. contentive), length in milliseconds and number of words preceding and following the target. They found that longer words (in terms of millisecond measurement), content words and words occurring later in the sentence tended to be recognised before acoustic offset. Shillcock et al.(1988) extended this earlier analysis by including some prosodic variables – right boundary strength (Gee and Grosjean (1983)), the presence of syllables with unreduced vowels and the position of such syllables in the word. Inclusion of these prosodic variables replaced the contentive/functor distinction as a predictive factor: instead, the presence or absence of unreduced syllables was seen to make a strong contribution, and right boundary strength played an important role.

Mehta and Cutler (1988) used regression to provide a more detailed analysis of their phoneme monitoring reaction times to read and spontaneous speech (2). The variables which they included in the regression were transitional probability, sentence position (counted from the first word in the sentence), sentence accent, vowel quality of the target-bearing syllable and preceding word length. They found that the effects of position in sentence, syllable (lexical) stress and sentence accent differed across speech modes: the effect of sentence position was significant in read but not in spontaneous speech, while the sentence accent and lexical stress variables were significant in spontaneous but not in read speech. These results were verified by a further series of analyses in which the continuous-scale variables were converted to binary classifications (in order to make the results comparable to those obtained in other studies); for example, transitional probability, which was originally conceived in their regression analysis as a score out of 20, was converted to high probability (higher than 5 out of 20 in the original analysis) vs low probability (lower than 6). These results corroborated those of the regression analysis.

Some of these variables were included in the regression analysis performed in the present study. However, it was not appropriate to include all of the variables employed by other writers. For example, the materials used in the experiments reported here are not amenable to the functor/contentive distinction included by Bard et al. The independent variables which were selected are discussed in detail below.

The set hierarchical regression method described by Cohen and Cohen (1983) was used to analyse the results of Experiment 3. In this technique, independent variables are grouped into sets according to logical criteria. Comparison of the multiple R-squared values which result from the inclusion of further sets in successive regression equations with those from the immediately preceding set yields a value of F which can then be used to determine the statistical significance of the contribution of the set of variables just included. If the difference between multiple R^2 values before and after the addition of a new set of variables is statistically significant, a t-statistic associated with the standardised regression coefficient for each variable can be assessed for significance. In addition, it is possible to examine subsets of the data separately and to determine the extent to which the independent variables contribute to the dependent variable in each subset: in the second analysis reported below, the read and spontaneous subsets are examined separately.

The data from Experiment 3 relating to the percentage of words recognised by acoustic offset were input to a regression analysis involving the word type variables, which were associated with the status of the stimulus as a word type, string variables, which were associated with the status of the stimulus as a member of a longer string of words, and word token variables, which were associated with the characteristics of the stimulus as it was uttered on a particular occasion. The variables included in each set were as follows:

WORD TYPE VARIABLES

Lexical Stress: words were either stressed or unstressed on the initial syllable.

Number of syllables: Bard et al. (1988) found that number of syllables was a highly significant contributor to the equation in the absence of the millisecond length variable. The number of syllables in the testwords in Experiment 3 ranged from two to four.

Word frequency: the frequency of each word (including frequencies for inflectionally related words) was determined using Francis and Kučera's 1982 word count.

Speaker: six speakers were represented in Experiment 3. It might be argued that Speaker should be included as a word token variable, because inter-speaker differences are commonly responsible for differences between word tokens (see Chapter 2). Since, however, word types were nested in Speaker in Experiment 3, it was decided to treat the latter as a word type variable.

STRING VARIABLES

Words before test word: the number of words which preceded the test word in the sentence which was presented to speakers for production in the read condition (see Chapter 5) was calculated.

Words after test word: the number of words following the testword was similarly calculated.

Cloze test result: transcripts of the sentence fragments preceding the testwords were presented to eight individuals who were asked to supply any word which could legally continue the fragment. These responses were then examined by a pair of judges who scored them according to the metric used by Marslen-Wilson and Welsh (1978): 4 for the testword itself, 3 for a synonym, 2 for a related word and 1 for an unrelated word. The mean of their agreed response scores for each testword was entered into the regression matrix: mean scores could thus logically range from 1 (all responses offered were unrelated to the testword) to 4 (all 8 subjects offered the test word itself). In practice, while many testwords were of low predictability, the highest mean transitional probability score attained was 3.2.

WORD TOKEN VARIABLES

Version: word tokens were assigned a code of 1 for stimuli excised from read speech and 2 for stimuli originally produced in conversations.

Length in milliseconds: the millisecond lengths of the test words (which were discussed in Chapter 5) were included in the analysis.

Sentence Accent: the read and spontaneous sentences containing the test words used in Experiment 3 were played to 4 listeners (postgraduates and fourth year undergraduates in the Linguistics Department of Edinburgh University) who were asked to mark a transcript to indicate which words they considered to bear sentence accent. The hearers were unaware which words had been experimental items. No definition of the term 'sentence accent' was offered by the Experimenter because it was felt that this would bias subjects' judgments. It was pointed out, however, that some items were drawn from conversational speech and that accentual patterns in such tokens might not conform to theoretical descriptions with which they were familiar. Each test word token was assigned a score ranging from 0 (= no listener judged the word to have been accented) to 4 (= all four listeners judged it to have been accented). In practice, there was a high degree of agreement between listeners with respect to the accentedness of items. Descriptive statistics for the accentedness judgments were as follows: Read/Syll-1: mean 3.14, s.d. 1.35; Read/Syll-2: mean 2.77, s.d. 1.51; Spontaneous/Syll-1: mean 3.08, s.d. 1.23; Spontaneous/Syll-2, mean 2.67, s.d. 1.60.

Initially it was intended that a further variable, vowel quality of the initial syllable, should be included. However, when the values for this variable were computed, it was discovered that, while all initially stressed tokens contained full vowels, only 3 words with initial unstressed syllables did so. It has been pointed out for some time by Cutler and her colleagues (e.g. Cutler and Clifton 1985; Cutler 1986) that studies purporting to examine the effects of stress on speech perception commonly confound this variable with vowel quality. The fact that the two variables were confounded in Experiment 3 no doubt reflects an important property of the variety of English found in conversational speech: that words with full, unstressed vowels in their initial syllables are not commonly found. Although there are many disyllabic words in English, the disyllabic words whose initial syllables both contain full vowels and are unstressed (e.g. *typhoon*) are relatively few in number: for example, Fudge (1984) includes only 77 monomorphemic words of this kind in a list purporting to be exhaustive. Furthermore, they seem to be generally of low frequency; of the words in this category listed by Fudge, 22 are too rare to appear in the Francis and Kučera (1983) word count, while of the remaining 55 words, 39 occur less than 10 times per million words of text. It is therefore unsurprising that so few items of this kind appeared in the sample selected for Experiment 3.

The 10 variables identified above were input to regression equations in which the dependent variable was the percentage of words recognised by acoustic offset. Results for the whole set of 192 word tokens are presented in Table 6.1., while Table 6.2. is concerned with the analyses of read and spontaneous tokens separately.

Table 6.1: Standardised regression coefficients for percentage of words recognised by acoustic offset, Experiment 3 (all tokens)

VARIABLES IN EQUATION	Type	Type & String	Type, String & Token
TYPE VARIABLES			
Stress	-.19**	-.154*	-.191**
Number of Syllables	.06	.109	.029
Frequency	.002	.008	.026
Speaker	.106	.081	.090
STRING VARIABLES			
Words Before		.212**	.227**
Words After		.013	.018
Cloze		.123	.127
TOKEN VARIABLES			
Version			-.330**
Length (msec)			.208**
Accent			.012
Y Intercept	83.93	58.45	77.34
Multiple R^2	.0467	.1062	.2773
F	2.291	3.124**	6.946**
d.f.	4,187	7,184	10,181
F for R^2 diff.		1.71	4.12**
d.f.	7,180	10,174	

* t-value significant at p < .05
** t-value significant at p < .01

When the data are considered as a whole, stress makes a significant contribution, which is unaffected by the introduction of subsequent sets of variables. The coefficient for stress is negative because the value of 1 indicated words stressed on the first syllable, which were associated with more recognitions by offset, while 2 indicated words stressed on the second syllable, which were less likely to be recognised by offset. The introduction of string variables into an equation containing type variables is relatively unimportant. However, when token variables enter the equation, the difference between multiple R^2 values is significant. Both version and millisecond length make significant contributions to the equation, but accent does not. The latter result is in conflict with the finding of Mehta and Cutler (1988). When the beta values for string and type variables are examined for their level of significance before and after the addition of token variables, it is apparent that their contribution to the equation does not change: stress and number of words preceding the test word are significant throughout.

Table 6.2: Standardised regression coefficients for percentage of words recognised by acoustic offset, Experiment 3 (read and spontaneous tokens analysed separately)

Notes:
F values for the addition of each new set of variables were less than 1 for the read sample and non-significant for the spontaneous sample.

VERSION	Read			Spontaneous		
VARIABLES IN EQUATION	**Type**	**Type & String**	**Type, String & Token**	**Type**	**Type & String**	**Type, String &Token**
TYPE						
Stress	.001	.018	.025	-.331**	-.277*	-.318**
Number of Syllables	.076	.059	.065	.099	.157	.038
Frequency	.034	.028	.026	-.014	-.002	.001
Speaker	.019	-.043	-.047	.195	.164	.143
STRING						
Words Before		.155	.153		.283**	.318**
Words After		.085	.085		-.023	.001
Cloze		.078	.077		.174	.200
TOKEN						
Length (msec)			-.018			.325**
Accent			.015			.005
Y Intercept	91.13	77.68	77.53	76.70	39.22	18.84
Multiple R^2	.0018	.0343	.0347	.1444	.2569	.3422
F	.041	.447	.343	3.838**	4.346**	4.972**
d.f.	4,91	7,88	9,86	4,91	7,88	9,86
F for R^2 diff.		.4039	.0032		1.82	1. [illegible]
d.f.		7,84	10,78		7,84	10,78

*** t-value significant at p < .05**
**** t-value significant at p < .01**

The comparison of the equations computed separately for read and spontaneous tokens (Table 6.2) reveals that the independent variables make different contributions in the two speech styles: the F value of 8.95 resulting from a comparison of the two conditions reached a high level of significance (d.f. = 5, 182; $p < 0.001$). When the two samples were assessed separately in this way, the inclusion of successive sets in the equation did not significantly affect its predictive power: in all cases, the value of F for the addition of further variables was close to 1.

The t-values associated with the standardised regression coefficients indicated the areas in which the two subsets differed. In the read condition, no variable contributed significantly to the equation. Part of the explanation for this no doubt lies in the comparatively high intelligibility of the test words themselves. It is clear that the significance of stress, number of preceding words, millisecond length and version itself when all words are examined as a single group (Table 6.1.) is attributable to their significance in the spontaneous condition.

To summarise, words which are initially stressed, occur late in a sentence, are relatively long or are uttered in read speech style have the greatest likelihood of being recognised by their acoustic offsets. Separate analyses of the read and spontaneous samples indicate that the contributions of stress, number of preceding words and millisecond length are significant in the latter but not in the former.

2. Use of context

A second issue which was examined in earlier chapters concerned the role of context in auditory word recognition. According to the Cohort Model, its use was confined to the retrieval stage; during access, only bottom-up information was taken into account. The Cohort Model defines this lexical access phase as corresponding to the initially 150 msec of a word (Marslen-Wilson and Tyler 1980; Marslen-Wilson 1987). The results of Experiment 1 contradicted this claim: subjects produced more contextually appropriate responses at early gates in the with-context conditions than in the no-context control condition.

In Chapter 4, the apparent failure of subjects in Experiment 1 to make use of contextual cues in some instances was discussed. Such failures could sometimes be attributed to the mis-analysis of preceding acoustic material (in particular, the mis-insertion of word boundaries). It was speculated in Chapter 4 that while this was a relatively unusual phenomenon in the carefully articulated materials of Experiment 1, subjects must not infrequently be misled by their misanalysis of preceding material in spontaneous speech if, as seemed likely, the level of intelligibility of spontaneous speech was poor. In the spontaneous conditions of Experiments 2 and 3, subjects frequently had difficulty in recognising words, suggesting that the overall level of intelligibility in these samples might be low. What, then, was the effect of this on their use of contextual cues?

An initial analysis was conducted to discover whether there was evidence that subjects in Experiment 3 had been using contextual cues in a similar manner in both version conditions. In chapter 4 it was shown that during the processing of the initial 150msec of stimuli from Experiment 1 (a portion which is assumed to correspond roughly to the lexical access phase) hearers made use of the contextual cues which were available to them. The same pattern emerged both when all the available data were analysed and when the analysis was confined to those responses offered by subjects upon hearing only the initial 50 msec. For the purposes of the analysis reported in this section, it was assumed that the processing of the read speech in Experiment 3 was similar to the processing of the speech in Experiment 1; that is, it was assumed that subjects processing read speech in Experiment 3 used any contextual cues which were available in accessing words. The comparison focused instead on the extent to which hearers also appeared to be using contextual cues when processing spontaneous stimuli in Experiment 3.

The responses offered by subjects after hearing the first gate in Experiment 3 were presented, along with a transcript of the sentence fragment which had preceded them, to 10 judges, who were asked to indicate which responses constituted valid continuations of the sentence fragment. The percentage of subjects who judged valid each response to each sentence fragment in each version condition was input to a 3-way (Version by Stress by Speaker) ANOVA. The results are presented in Tables 6.3.

and 6.4. (means collapsed across Speaker). While the Stress and Speaker effects and the interactions failed to reach significance, the Version means were markedly different: when subjects in the gating experiment had heard a spontaneous sentence fragment followed by a test word, it was far less likely that their response would be appropriate to context than when they heard a test word with the same lexical characteristics (i.e. another token of the same word type) preceded by a read speech sentence fragment.

Table 6.3. Percentage of responses from gate 1 of Experiment 3 which were judged appropriate continuations of preceding sentence fragments

VERSION	Read	Spontaneous	Mean
STRESS			
Syll-1	62	46	54
Syll-2	64	44	54
Mean	63	45	

Table 6.4: ANOVA for data in Table 6.3, related to Version, Stress and Speaker

Source	F
Version	(1,84) 24.36 **
Stress	(1,84) 0.00 n.s.
Speaker	(5,84) 0.82 n.s.

Although the read mean was somewhat lower than the with-context means reported in Experiment 1, this is no doubt due in part to the fact that the strings used in the experiment were naturally occurring utterances which could well be less predictive than the designed sentence fragments of Experiment 1.

The only main effect or interaction which reached significance was Version. The likelihood of a response's being appropriate to context was markedly higher in the read than in the spontaneous condition. The results of Experiment 1 indicated that hearers made use of any contextual cues which were available to them, and this also appears to be the case for the read speech sample in Experiment 3. Why should they fail to do so when processing the spontaneous speech sample? A likely explanation (or at least part of an explanation) seems to be that the words containing the contextual cues are often, like the testwords themselves, of poor intelligibility. To test this hypothesis, a subset of the materials from Experiment 3 were further analysed.

2.1. Experiment 4

2.1.1. Method

2.1.1.1. Materials

The results of the semantic appropriateness analysis just described were examined, and for each speaker, four tokens were selected (representing two stimulus types, i.e. the read and spontaneous versions of two different stimuli). The stimulus types which were selected were those for which there was the largest discrepancy between the read and spontaneous score in each Speaker X Stress cell in the analysis just described. This procedure resulted in the selection of 24 stimuli overall. The stimuli, up to but excluding the test word, were recorded onto audio tape to be presented to subjects in an intelligibility test. The mean length of the sentence fragments thus produced was 5.9 words; this figure was of course identical across the two version conditions. Two tapes were constructed, each containing 12 stimuli. If a stimulus appeared in its read version on one tape it appeared in its spontaneous version on the other tape. Speakers and stress levels were evenly represented on the two tapes. Two repetitions of each stimulus were recorded, each preceded by a warning tone.

2.1.1.2. Subjects

Each tape was presented to two different groups of seven listeners, undergraduates or postgraduates in the department of linguistics.

2.1.1.3. Procedure

Subjects were asked to provide a word level transcription of what they heard. They were warned that the tapes contained spontaneous tokens which might prove difficult to understand, but they were encouraged to provide some response in each instance.

2.1.2. Results

The responses of the listeners were examined to determine how many words, in the correct order, they had identified correctly. The mean number of words correctly identified was computed for each utterance token, and the results input to a matched t-test. The difference between the accuracy of recognition across the two version conditions was highly significant ($t(1,11) = 6.52$, $p < 0.001$, by materials): while an average of 3.12 (53%) of the read words were recognised, only 1.21 (21%) spontaneous words were recognised.

Thus it appears that at least part of the reason that the responses of subjects in the spontaneous condition of Experiment 3 were less appropriate to context than those offered when subjects heard the corresponding read stimuli was that in the former condition they often failed to identify the words which composed the left context of the testword. A more detailed explanation of the circumstances is perhaps desirable. For example, do hearers recognise some of the words comprising the left context and attempt to construct a contextual framework from those words? Such a strategy might be reasonably successful for much of the time. Bard et al.(1988) have shown that there is a greater likelihood that content words will be recognised by acoustic offset than function words, a result which is mediated by the tendency of content words to be of longer duration than functors, and by the tendency of content words to contain a strong syllable. Since contentives are generally assumed to carry a greater information content than functors, it ought to be possible to derive relatively reliable semantic cues from these. The cues provided by functors, by contrast, will more often relate to syntactic category, which is comparatively ineffective at narrowing down a class of word candidates (Tyler and Wessels 1983). However, when the results of Experiment 4 were analysed with respect to the functor/contentive distinction, the results did not support this interpretation. The cell means presented in Table 6.5 show a version effect which is highly significant (see Table 6.6); but the functor/contentive distinction itself, and the interaction between the two variables, did not exhibit a reliable difference. However, these results were based on a very small sample; a larger study might indicate a relationship between the recognition of contentives and that of subsequent words.

Table 6.5: Percentage of functors and contentives correctly recognised in sentence fragments preceding test words (subset of Experiment 3 materials)

FORM CLASS	Functor	Contentive	Mean
VERSION			
Read	87	93	90
Spontaneous	23	25	24
Mean	55	59	

Table 6.6: ANOVA for data in Table 6.5.

Source	F1	F2	MinF′
Version	(1,13) 186.97 **	(1,19) 114.95 **	(1,31) 71.19 **
Form Class	n.s.	n.s.	---
Version X Form Class	n.s.	n.s.	---

* $p < .05$
** $p < .01$

A further aspect of the use of context which requires examination is whether hearers in fact *attempt* to use contextual cues when they process spontaneous speech. It has been shown that hearers can tell from the prosodic structure of a speech excerpt whether it was produced spontaneously or read from a script (Johns-Lewis 1987). Perhaps hearers of spontaneous speech, realising that the contextual cues which they perceive may be unreliable because of their misperception of words, do not attempt to use contextual cues. This account seems implausible, however: since individual spontaneous word tokens are less intelligible than those in read speech, it seems unlikely that hearers would ignore a potentially useful source of information such as context in this way; an analogy might be drawn with the use made of context by inexperienced readers to compensate for their poor bottom-up processing skills (Ellis and Beattie 1986). It seems more likely that hearers do attempt to use contextual cues as long as any are present and can be perceived, but that they must relatively often be misled by this strategy. No doubt the computational consequences of this approach are costly, but this is consistent with the evidence of this study which indicates that conversational speech is on all measures more difficult to process than read speech.

3. Phonological access

Models of word recognition differ with respect to the part of the word which is used to access representations in the mental lexicon. The candidates which were considered in Chapters 4 and 5 were initial syllables on the one hand and stressed syllables on the other. Several analyses suggested that the models which used initial syllables as the key to the mental lexicon could not accommodate the findings of this study with respect to the stress variable.

Cutler and Clifton (1984) and Cutler (1986) have pointed out that experiments purporting to examine stress differences (e.g. Bond 1971; Cole, Jakimik and Cooper 1978; Shields, McHugh and Martin 1974; Cutler and Foss 1977) in fact commonly confound stress with vowel quality. Within the stressed/unstressed distinction, two parameters are available for manipulation: on the one hand, the phonological factor of vowel quality (full versus reduced) and on the other, phonetic factors such as duration, amplitude and fundamental frequency movement. It was noted in Chapter 2 that the former is in some sense the more robust feature; stressed syllables may be realised with none of the phonetic characteristics, but the vowel quality will tend to be preserved to some extent.

Altmann and Carter (in press) have suggested that it is the phonological attribute of vowel quality which makes stressed syllables more informative than unstressed syllables. They discovered that the wider range of vowels which occurred in stressed syllables contributed to their creation of lower entropy (i.e. higher informativeness). The belief that the vowel quality parameter ought to contribute to informativeness was, indeed, one of the characteristics of stressed syllables which made it seem an attractive candidate as the key to the mental lexicon (see Chapter 2).

Altmann and Carter's study was based purely on segmental transcriptions and therefore ignored the phonetic aspects of stressed syllables. However, it was argued in Chapter 2 that these phonetic characteristics might also increase the effectiveness of stressed syllables as input units which could trigger lexical access: it should be possible to identify fundamental frequency movement and increased amplitude in the acoustic signal (thereby giving the hearer some orientation with respect to the position of acoustic material in prospective word candidates), and increased duration and amplitude seem likely to enhance perception. Cutler and Norris (1988) emphasise the phonetic characteristics of speech perception: they have claimed that hearers employ a segmentation strategy which involves attention to durational and other aspects of speech (though only as a means of identifying the quality of a vowel, by making use of a likely correlation between performance and linguistic phenomena).

Thus, it might be argued that either the phonetic or the phonological characteristics of stressed syllables account for their effectiveness as keys to the mental lexicon. Further analysis of the data from this study is required before this issue can be resolved. In Section 6.1. above, it was noted that the stress and vowel quality variables had been confounded in Experiment 3. Furthermore, the regression analysis reported in that section revealed that the accentedness of a word token did not contribute significantly to its likelihood of being recognised. Since accent is a suprasegmental phenomenon with phonetic consequences, it seems likely that the aspect of the words stressed on the second syllable which led to their later recognition in Experiment 3 (in the spontaneous condition at least) was not stress *per se* but the related, and in this experiment uncontrolled, variable of vowel quality.

Unfortunately, because the two factors were confounded in Experiment 3 it is not possible to examine this issue directly using data from that experiment. However, in

Experiment 1, the words stressed on the second syllable were reasonably evenly divided between those with full initial syllables and those with reduced initial syllables. Of these 24 words, 9 had initial syllables with full vowels while the remaining 15 had reduced vowels in their first syllables. A two-way ANOVA (syllable structure (2 levels: full vs reduced) X context (3 levels: none, short and long)) was used to determine whether there was evidence for differential processing of words with reduced and unreduced initial syllables. The dependent variable which was used was the number of subjects, out of a total of 12, who, when presented with the initial C(C)V of the stimulus, had produced a response which matched the stimulus for initial C(C)V. (See Chapter 4 for an account of the rationale underlying this analysis.) Stress is held constant in this analysis, since the stimuli which subjects heard when they produced the responses being analysed were all unstressed initial syllables. The crucial difference between the syllabic stimuli was that in some the vowel was reduced (as in the word *computer*) while in others it was full (as in the word *cartoonist*). If the segmental characteristics of stressed syllables are important, a significant difference should be found between the two groups of stimuli. Results are presented in Tables 6.7 and 6.8.

Table 6.7: Mean number of hearers (N = 12) who offered responses consistent with a C(C)V stimulus (Experiment 1: words with second syllable stress only)

CONTEXT	None	Short	Long	Mean
INITIAL SYLLABLE				
Full	8.4	8.7	9.8	9.0
Reduced	3.6	5.6	8.5	5.9
Mean	5.4	6.8	9.0	

Table 6.8: ANOVA for data in Table 6.7

Source		F	
Syllable structure	(1,22)	13.66	**
Context	(2,44)	7.63	**
Syllable Structure X Context	(2,44)	2.41	n.s.
* p < .05	** p < .01		

The results support the hypothesis that the full/reduced distinction within unstressed vowels is extremely important. However, no conclusions can be drawn from this analysis about the importance of phonetic characteristics such as amplitude and duration. To examine this issue, another comparison can be made among the materials and data from Experiment 1. The nine stimuli in the previous analysis which had full vowels had partners, in the set of words stressed on the first syllable, which matched them for at least the initial C(C)V. If these nine pairs, which matched for segmental characteristics in the initial portion, differ across stress groups in terms of subjects' ability to access a set of words consistent with the initial C(C)V stimulus, then this will be evidence of the importance of phonetic factors such as amplitude and duration, since phonological characteristics relating to vowel quality will have been held constant. The results of the analyses are presented in Tables 6.7. and 6.8.

Table 6.9: Mean number of hearers (N = 12) who offered responses consistent with a C(C)V stimulus (Experiment 1: words with full vowels in initial syllable, matched pairs from both stress conditions)

CONTEXT	None	Short	Long	Mean
STRESS				
Syll-1	6.7	10.3	10.9	9.3
Syll-2	8.4	8.7	9.8	9.0
Mean	7.6	9.5	10.4	

Table 6.10: ANOVA for data in Table 6.9

Source		F	
Stress	(1,8)	0.33	n.s.
Context	(2,16)	5.22	*
Syllable Structure X Context	(2,16)	3.27	n.s.
* $p < .05$ ** $p < .01$			

The stress difference was insignificant overall, which suggests that phonetic factors played no role. Indeed, in the No Context condition the predicted direction of the difference is reversed: words stressed on the second syllable elicit a greater number of matching responses than words stressed on the first syllable. A planned comparison of means revealed no reliable difference between the stress levels in any of the three context conditions. The evidence of this analysis is thus that the phonological characteristics of stressed syllables are more important, in terms of their role in word recognition, than phonetic characteristics. Note, however, that this conclusion is based on an analysis of performance on carefully articulated materials. If it were possible to conduct a similar analysis using materials drawn from a spontaneous sample, the phonetic characteristics of stressed syllables might prove an important factor.

The evidence of this study suggests that stressed syllables play an important role in lexical access and that this importance is attributable to the phonological properties of the syllables. However, the issue of how this behaviour is to be encoded in a model of word recognition remains open. Although the analyses presented in Chapters 4 and 5 indicate that stressed syllables provide useful information to the processing mechanism, this might not be because of their status as stressed syllables *per se* but because of their relative information value. Altmann and Carter have shown that this arises in part because of the segmental characteristics of stressed syllables and analysis of the data from Experiment 1 has shown that when phonetic differences associated with stress are controlled, the vowel quality differences between full and reduced syllables can still be used by hearers. One method of encoding this ability might be to propose a model in which lexical access proceeds via stressed syllables. But it is not only in stressed syllables that full vowels are found: some unstressed syllables may be protected from vowel reduction, depending on their phonological composition (Fudge 1984).

Cutler and Norris (1988) have proposed a model in which full vowels cause the word recognition device to insert tentative word boundaries before the syllables containing them. Such a mechanism might explain some of the findings of this study, but it would fail to account for others. Experiments 2 and 3 revealed large differences between read and spontaneous speech as well as between stressed and unstressed syllables. If either the stressed syllable notion or the strong syllable segmentation strategy is incorporated into the model to account for the effects relating to stress/vowel quality, then a separate mechanism must be built in to account for the read/spontaneous differences. A more parsimonious account would use a single mechanism to account for both sources of variability, which, as we saw in Chapter 2, affects both vowels and consonants in the read/spontaneous dichotomy.

In order to account for all of the results reported in this study, a model would need to be able to accomplish lexical access and lexical retrieval despite input of differing quality; but the performance of the model would need to reflect the fact that poor quality input (e.g. unstressed syllables, spontaneous word tokens) is processed less efficiently by hearers than good quality input (e.g. stressed syllables, read word tokens). A number of recently-proposed word recognition models seem capable, in principle, of achieving this: Marslen-Wilson's (1987) modified Cohort Model; McClelland and Elman's (1986) TRACE model; and various attempts at automatic word recognition using the Hidden Markov Model (HMM) technology (see, for example, Bridle et al. 1982; Rabiner and Juang 1986; Cox 1988). What these models have in common is that input is analysed in featural terms and a goodness-of fit metric is used to map onto lexical entries similarly specified; word candidates are then activated to varying degrees depending on the extent of the match between input and stored representation. In this approach, the activation process corresponds to lexical access; but the phenomenon responsible for lexical retrieval is typically underspecified. It is generally assumed that lexical retrieval will occur when the level of activation of the winning candidate reaches a criterial level or when activation of one candidate outstrips that of all others.Whatever method is adopted, recognition proceeds as follows. When the system encounters a relatively 'good' token of a word (or syllable), as might be the case when a read rather than a spontaneous word, or a stressed rather than unstressed syllable is presented, activation will be high and the word will be more easily recognised. However, when the processor encounters a poor token (unstressed syllables, spontaneous utterances) activation will be low and recognition will consequently be delayed until enough information has accumulated to allow the activation level to reach threshold. The notion of 'threshold' has generally been left undefined in theories which make reference to it: for example, it could have some absolute value, or it might reflect some differential between the activation levels of the top two candidates, or between the top candidate and all the others. Bard (forthcoming) discusses the consequences of each of these interpretations.

Although Marslen-Wilson (1987:78) continues to suggest that the more recent version of the Cohort Model relies on the importance of word onsets as the means of entering the mental lexicon, this is actually an unnecessary feature of the model if the activation notion is incorporated. At certain points in the speech stream the processor will indeed 'know', or at least feel fairly confident, that a word boundary is present. Such points might be located when an acoustic word boundary cue is encountered, for example, or when the hearer has recognised preceding words with a suitably high level of confidence, or when the speaker has started speaking after pausing. Note that the first two sources of information about word boundaries are very often available in read speech, as this study has shown, while the third is more common in spontaneous utterances (see, for example, Mehta and Cutler 1988). When the hearer is reasonably confident about the location of word boundaries, for whatever reason, processing of the next word can begin in a left-to-right fashion, and if the quality of the acoustic material is sufficiently high then this word, too, will be recognised by offset and will indicate the location of the subsequent word. This account, which assumes the availability of word boundary cues from some source and the high intelligibility of individual words, is consistent with the state of affairs in carefully articulated connected speech such as was used in Experiment 1 and in the majority of word recognition studies reported in the literature. It is not at all surprising, therefore, that experimental results have failed to challenge the prevalent view of word recognition as a word-by-word, left-to-right phenomenon. The only evidence of differential intelligibility is at the intra-word level: when the early parts of words have unstressed syllables (are relatively unintelligible) the set of word candidates which is activated is large (that is, responses tend to match poorly with the input because of the poor quality of the token, and therefore a larger set of candidates is accessed). It is only at a later stage, when comparatively good quality bottom-up information is coupled with top-down cues, that the set of word candidates can be narrowed down and recognition can occur; but in highly intelligible speech, this has almost always happened by the ends of words, both because word tokens themselves are highly intelligible and because the context, which is also more intelligible, can be used. Therefore, when the dependent variable is a measure of lexical retrieval (or 'selection', according to Marslen-Wilson (1987)), rather than a measure of lexical access, the effects of different stress patterns are not discernible. But when lexical access itself is examined, stress effects should be found. This explanation is consistent with the findings of Experiment 1.

However, when we turn our attention to spontaneous speech, or indeed to any form of speech which is marginally unintelligible (like, for example, the read sample in Experiments 2 and 3), a very different picture emerges. The opportunities for identifying word boundaries are limited; the only cues which are available on a regular basis are pauses. (Note that the occurrence of phonological phrase boundaries correlates well with pause duration (Gee and Grosjean 1983), and that when hearers encounter phonological phrase boundaries, recognition of previously unrecognised words often occurs (Shillcock et al. 1988).) When pauses are encountered, the processor can be re-started, as it were, and if the material following the pause is relatively well-defined, word-by-word recognition may be observed. However, if poor bottom-up information is coupled with an inability on the part of the hearer to decipher the left context of the string (such as was shown to be the case for the spontaneous sample in

Experiment 3), it is unlikely that word recognition of the current token can be accomplished efficiently. Large numbers of temporally overlapping candidates are activated, and the resulting set is only narrowed down when high-quality material (such as that found in stressed syllables) is encountered. Under such circumstances, the superiority of informative (e.g. stressed) syllables which was found at the intra-word level in carefully articulated speech is likely to persist and have consequences at the retrieval stage, with the stress difference reaching significance for such dependent variables as the percentage of words identified by offset and the percentage of the word heard before identification occurs. This account is consistent with the findings of Experiments 2 and 3. For this explanation to be viable, it is necessary to propose a second lexical retrieval mechanism apart from those mentioned above (attainment of an activation threshold or of activation differential compared with other candidates). In cases where several words have elapsed and no candidate has been retrieved by either of these methods, the occurrence of a particular event in the speech stream, such as a pause, would force the processor to recognise the (syntactically well-formed) set of words with the highest overall activation level. The fact that top-down information can contribute to activation would tend to result in the recognition of strings of words which were appropriate to context.

It was noted that the read sample in Experiment 2, while more intelligible than the spontaneous sample in the same experiment, was markedly less intelligible than the sample used in Experiment 1. When the tokens from read and spontaneous speech, stressed on the first and second syllables, were presented in isolation, the version difference and the stress difference, but not the interaction between these variables, reached significance for dependent variables relating to both the access and retrieval stages. But when, in Experiment 3, context was available to hearers, a Stress by Version interaction emerged. At the earliest stages of processing, a weak stress effect was found; but when the dependent variable related to retrieval (number of isolations, percentage of word heard before isolation) the stress difference was found in the spontaneous sample only. In this chapter, evidence has been presented which indicates that the left contexts of read tokens were more intelligible than those of spontaneous tokens. Thus, when processing words in the read sample, subjects had the assistance of contextual cues (which, as Experiment 1 showed, were indeed used at the access stage). The presence of such cues tended to remove the effects of stress when retrieval measures were considered. However, in the spontaneous condition, contextual cues were of little assistance to hearers because they were so unintelligible, and accordingly the stress effect persisted.

Chapter Notes

(1) Homophonous items (e.g. *bare* and *bear*) are an obvious exception to this claim; if the information which identifies the required interpretation occurs after the homophonous item, the word sense cannot be identified until than information has been perceived by the hearer.

(2) Mehta and Cutler were not explicitly interested in the factors which enabled hearers to *recognise* words by their offsets, but in the speed with which they could respond to phoneme monitoring targets. However, since the effects of transitional probability of the target-bearing word were found to be significant for both speech modes, it seems likely that their subjects were using the lexical rather than the phonetic route to phoneme monitoring responses (Foss, Harwood and Blank 1980).

CHAPTER 7

Conclusion

This study set out to examine three issues related to the question of the way in which words are recognised in fluent speech: first, the extent to which hearers can rely on the recognition of words before their acoustic offsets; second, the role of context in lexical access; and third, the phonological unit which is used to access the mental lexicon. All three issues have been considered with respect to read and spontaneous speech.

The first aspect of the word recognition process which must be considered in the light of the experimental evidence presented in this study is the late recognition of words in spontaneous speech. The consequence of a substantial proportion of failures to identify words by their offsets would be that this source of information about the location of word boundaries would be unavailable to the hearer. Since it is well-known that phonological word boundary markers are often absent from casual speech, this failure is potentially serious, particularly for models of the word-recognition process which emphasise the importance of processing the beginnings of words. The results of Experiment 1, which assessed the intelligibility of the kinds of speech materials which are commonly used in word recognition experiments, were highly consistent with the requirements of such models, since all but a few words were identified even when presented in isolation. When different speech styles (spontaneous speech, and speech read from a transcript of a conversation) were examined, however, the results were incompatible with such an account. Particularly in spontaneous speech, large numbers of words could not be recognised by offset, even in the presence of a substantial portion of context. Indeed, it was sometimes only after one or more subsequent words had been heard that recognition of test items occurred. Other studies have pointed to the unintelligibility of individual words in spontaneous speech. The results of Experiments 2 and 3 have indicated that the failure to recognise words by their acoustic offsets is a factor even in the processing of relatively long content words, which other research reported in the literature might lead us to suppose would be most likely to be recognised on time.

The fact that a substantial proportion of content words are recognised late in spontaneous speech has consequences for the second aspect of processing, the use of contextual information. While some models restrict the use of contextual cues to particular phases of processing, the results of Experiment 1 indicate that hearers use top-down information from the earliest point in their processing of speech: gating responses in the with-context conditions were clearly influenced by syntactic and semantic considerations, compared with the responses to identical tokens in the no-context condition. Therefore, restrictions on the use of contextual cues, such as those imposed in the Cohort Model, must be lifted. Certainly it has not been possible to identify, in this study, a period during the processing of read words from which contextual influences were absent. If the notion of strong interaction between components is to be rejected, as several authors have argued it should be (e.g. Simon 1969; Crain and Steedman 1985), and a weakly interactive model adopted, the point at which the top-down information is allowed to be used is much shorter than the 150 msec estimated by Marslen-Wilson (1987); in fact, it is even shorter than 50 msec, according to the evidence of this study. There comes a point at which, if a weakly interactive model continues to be proposed, the period of bottom-up autonomy becomes so short that weak interaction becomes as untestable as strong interaction is claimed to be.

This study has revealed effects associated with top-down processing at very early stages in word recognition. Yet contextual information cannot solve all the ills of the auditory word recognition process. In order to apprehend syntactic and semantic cues hearers must recognise at least some of the words of which they are composed. For example, when subjects in Experiment 1 heard the sentence fragment

(7.1) When I went to see the doctor, he was in the ...

the presence of the lexical item *doctor* would contribute to their processing of the test word *surgery* if and only if they were able to recognise it prior to their processing of *surgery*. Similar comments apply to the utility of syntactic and semantic cues. It may often be the case that hearers are unable to utilise recent contextual cues efficiently in casual speech. If the acoustic material containing the contextual information is of low intelligibility, the cues which the hearer derives will be at worst misleading and at best non-existent. Post-hoc analysis of the data and materials from Experiment 3 provide relevant evidence. When the responses offered to read and spontaneous stimuli were assessed for the degree to which they were consistent with context, it was clear that subjects were less efficient at using contextual cues in the spontaneous mode, since their responses were judged to be less appropriate to context. When a subset of the stimuli were later examined to determine the intelligibility of the left contexts of the test words, it was found that spontaneous sentence fragments were far less intelligible than read ones. A possible explanation, then, of hearers' apparent failure to use contextual cues when processing spontaneous word tokens was that they had failed to process accurately the acoustic material containing the cues.

A similar explanation may perhaps account for the interaction between context and literacy in Experiment 1. On all the measures adopted, subjects in the low literacy group were shown to be less efficient at using context than their high literacy counterparts. It cannot be argued that this interaction was attributable to some aspect of the predictability of the test words, because separate cloze tests were conducted using subjects from both populations prior to the auditory experiment, and the transitional probability of the words was equal for both groups. Instead, it seems more likely that the low literacy group had problems either with the auditory processing of the acoustic material preceding the test words or with integrating the semantic representation of recognised words into the contextual framework and using that framework to assist the lexical access procedure when processing the test words themselves. The reason for replicating Experiment 1 with the low literacy subjects was in fact to test the hypothesis that they would be less efficient at bottom-up processing than the undergraduate group, due to less highly developed phonological mapping skills. While there is no direct evidence of this from the data of Experiment 1, it is certainly possible that this factor is implicated in the context by literacy interaction: that is, perhaps the reason the Low Literacy group used contextual information less efficiently than the high literacy group arose from problems with their recognition of words preceding the test items, problems which might themselves be attributable to impaired phonological processing.

The third issue which was examined once again concerned processing activity during the lexical access phase. The original version of the Cohort Model, which suggested that lexical access depended on hearers matching the initial portions of input and lexical representation at a phonological level, ignored the possibility of initial syllables of differing intelligibility, as well as being susceptible to problems arising from the lack of segmentation cues already discussed. That initial syllables do differ with respect to their efficiency as a means of accessing lexical candidates is evidenced by the results of all three experiments.

In this study, two sources of variability have been identified which affect the utility of initial syllables as phonological cues to the identity of words. First, an unstressed syllable is a less efficient means of accessing the mental lexicon than a stressed syllable. Even in Experiment 1, in which the materials were clearly articulated, stress effects were found during the access phase of processing. However, by the retrieval phase of processing, which was investigated in this study by use of the isolation measure, stress effects were no longer discernible.

Given the presence of the stress effect at the access stage, it is tempting to suggest that it is stressed syllables which cue phonological access. However, such an account is not entirely acceptable because it is not only stress level which has consequences at the access stage. Experiments 2 and 3 indicated the influence of speech style: when read and spontaneous speech are compared, read tokens are more effective initiators of lexical access, in terms of extracting a set of phonologically consistent word candidates, than spontaneous tokens. If the stressed syllable model of access is adopted, a further mechanism is required to explain the version differences which were found.

A more parsimonious account would explain both the stress and version effects using a single mechanism. Such a mechanism might be the kind of activation-based process proposed by Marslen-Wilson (1987), McClelland and Elman (1976) and others. Marslen-Wilson suggests that hearers locate the beginning of a word and derive some low-level (e.g. featural) representation of the input starting at this point. If the stimulus is relatively indeterminate (e.g. an unstressed syllable, a spontaneous word token, or -- even more so -- an unstressed syllable in a spontaneous word token) this will result in the activation of a large set of word candidates. Poorer articulation leads to the identification of fewer features, and if the featural description of the input is under-represented, a large candidate set is identified. Subsequent acoustic-phonetic material, as well as preceding context, will, however, narrow down this cohort and result in retrieval, according to Marslen-Wilson, by word offset in most cases.

The interesting aspect of this account is that 'better' tokens (e.g. stressed syllables, or carefully articulated read words) will access fewer lexical candidates on the one hand and will be more successful at narrowing down a large cohort than poorly specified tokens. It is this characteristic that enables this proposal to account for the stress effect. When the hearer is confronted with a word with an unstressed initial syllable, a large number of candidates is activated, though to a low level. When the stressed syllable is perceived, this large set is quickly narrowed down to a small number, but because there is necessarily some time delay in the arrival of the stressed syllable, 'cohort entry', as it has been termed in earlier chapters, is relatively delayed in words with unstressed initial syllables. In a word with initial stress, however, the good-quality acoustic information is available at the start of the word, so a smaller number of candidates, which are likely to match the input fairly well in phonological terms, are activated. Thus the cohort entry point will be comparatively early.

Although this basic mechanism can account for the results reported here, it must be modified in a number of ways to explain other effects described in the literature. First, it has been repeatedly pointed out in earlier chapters that the flaw in the Cohort Model's account of word recognition is its insistence that hearers should be able to locate word onsets on a regular basis. However, with the computational technique just outlined such a requirement is no longer necessary. Suppose that some event indicates to the hearer that a word boundary is located at a specific point, which we will call t_1. Such an event might be the occurrence of a pause, or a word boundary marker of the kind discussed in Chapter 2, or the recognition, with sufficient confidence, of the preceding word. At point t_1, then, a number of lexical entries are accessed. Similar access attempts are made at points t_2, t_3 etc, but assuming that the acoustic material being analysed is indeterminate in quality, the large number of hypotheses which must be accessed reach only a low level of activation. At point t_n, however, some clearly articulated portion of speech is encountered. The existing pool of candidates is assessed to determine whether any of the activated candidates contain the clearly articulated material. This search may reveal a candidate extending all the way from point t_1 to the clearly articulated material, but in fact any candidates which are consistent with the well-specified acoustic material will automatically have their activations boosted while those of the other members of the candidate pool will decrease. If a candidate which does not span the whole distance from t_1 to t_n is among the set which is highly activated, an attempt will be made to find a plausible word or string of words from among those activated which are consistent with the material which the hypothesis does not span. If recognition proceeds in this manner, it is unnecessary to postulate the availability of explicit word boundary information for each and every word; such information would be used if it were present, but otherwise the identification of word boundaries is a by-product of the recognition technique. A similar account has been proposed by McClelland and Elman (1986).

Marslen-Wilson's 1987 model requires further modification if it is to explain the results of Cutler and Norris (1988), who found that recognition of a real word which straddled the two syllables of a non-word was delayed if the second syllable contained a full vowel, relative to the condition in which the vowel in the second syllable was reduced. For this phenomenon to be explained, the model needs to allow lateral inhibition, so that when two word candidates overlap, each suppresses the activation of the other. If the later candidate contains a full vowel it will tend to be strongly activated and will therefore either inhibit the earlier candidate, or delay its recognition, to a greater extent than if it contains a reduced vowel. Such an explanation of Cutler and Norris' results has been proposed by Bard (in preparation). The TRACE model (McClelland and Elman 1986) is similar to the model proposed by Marslen-Wilson (1987), except that it incorporates a lateral inhibition mechanism of the kind just discussed.

It might reasonably be objected that such an approach would quickly lead to a computational explosion of massive proportions. However, this is so only if *no* word boundary is recognised with reasonable confidence for a stretch of perhaps nine or ten words. However, this condition is rarely met in any speech style. Consider first the case of the perception of words in carefully articulated, isolated sentences, such as one might encounter in a typical psycholinguistics experiment. Except in the case of disfluencies (2), the hearer knows that the onset of a sentence is also the onset of a word. The processor begins to activate word candidates on the basis of the first portion of acoustic material. If this is a relatively carefully specified stretch of speech such as a stressed syllable, the number of candidates activated will be relatively small, and their activation level relatively high. Subsequent acoustic material in the word will often result in its being retrieved by its offset (retrieval occurring when the activation threshold is reached), and if so, the processor can begin afresh with the second word, reasonably confident that it begins where the previously recognised word ends. Furthermore, the presence of context will result in the high activation of only a few candidates at the access stage; this will also tend to lead to recognitions by offset.

Occasionally, however, even in the most carefully articulated speech, hearers do not recognise a word before its offset, and the pool of candidates must persist beyond the (actual, though not perceived) word boundary. This was the state of affairs reported by Grosjean (1985), in his gating study of the processing of low frequency monosyllabic content words in standard laboratory-produced (i.e. read) materials. Two factors must be taken into account, however. The first is that, in carefully articulated speech, recognition of word 1 may not be the only means of locating the onset of word 2; there may be explicit segmentation cues in the speech stream. The second point is that, even if the indeterminacy persists, it is unlikely to do so for more than a few words. For example, in Grosjean's study, almost all recognitions were achieved within about four words, and most occurred within the subsequent word. Thus, in read speech, the computational explosion will be confined to a span of around four words, which should not be beyond the capacity of a sophisticated parallel architecture (3). Very often, however, recognition will be accomplished in a shorter time span. To summarise, this is made possible by the general intelligibility of read speech on the one hand and the availability of word boundary cues (both explicitly marked and inferred from previous recognitions).

Consider now the state of affairs in spontaneous speech. Is it possible to suggest that the processor can expect indeterminacy to be confined to a span of only four or five words? Fortunately, the answer appears to be yes. Although the sources of word boundary information which are available in read speech -- phonologically marked boundaries, and the possibility of inferring the presence of boundaries through efficient recognition of preceding words -- cannot be counted on in casual speech, a further source of this information should not be overlooked. It has been noted in other studies (e.g. Mehta and Cutler 1988) that speakers pause more often in casual than in formal (e.g. rehearsed or read) speech. The result of this frequent pausing is that after only a few words the hearer will have a clear indication that a word offset has been encountered, and a word onset will shortly appear. (Once again, this account deliberately ignores the possibility of disfluencies.) Within this pause-defined, possibly multi-word chunk, processing can proceed much as outlined above, save that the degree of indeterminacy will be greater because of the less precise articulation in casual speech. Note, however, that the role of stressed syllables will be just as important as in read speech, if not more so, because they will continue to provide high-quality information. If a pause is encountered before any candidate has been sufficiently activated to reach threshold, the processor uses an alternative retrieval mechanism to select the word string to be recognised: those words with the highest overall level of activation are retrieved.

This account, in which pauses are used to define processing chunks, is consistent with that proposed by Shillcock et al. (1988) who report experimental evidence suggesting that hearers use prosodic cues to segment the input into multi-word sequences. Interestingly, the phi algorithm they used in their study (Gee and Grosjean 1983) also predicts more pauses at locations of higher right boundary strength. Shillcock et al.(1988) found that previously unrecognised words were often recognised as a word of high right boundary strength was presented.

When words of comparatively low intelligibility are excised from context (for example, in both the read and spontaneous conditions of Experiment 2) the stress effect will be apparent at both the access and retrieval stages: at the access stage for the reasons outlined above, but at the retrieval stage also because the low level activation of a large pool of candidates early on in the word (which is the case for words with unstressed initial syllables) is likely to result in a longer time being needed to resolve the cohort, while a good-quality syllable at the beginning of a stretch of speech (e.g. a stressed word-initial syllable) results in high activation of few candidates, and can thus be relatively quickly resolved. Some attempts at automatic speech recognition have adopted approaches similar to the one just outlined (see for example Bridle et al.1982; Cox 1988)).

One further point must be made in order to explain the version by stress interaction which was found in Experiment 3. Although stress effects were found during the access phase for both read and spontaneous tokens, by the retrieval stage, the effect was absent from the read samples (both in Experiment 1 and in Experiment 3) but persisted in the spontaneous speech. This interaction can be simply explained in terms of the use of context to constrain the access procedure. In Experiment 3, the sentence fragments which preceded the read words were more intelligible than those which preceded the spontaneous tokens, with the result that the responses that subjects offered during the access phase were more likely to be contextually appropriate in the read than in the spontaneous condition (4). So in the read condition, hearers would begin their processing of the test words by accessing a comparatively small set of candidates, whether or not they had good quality acoustic material at the start of the word. The facilitating effect of the position of high-quality acoustic information in the word was thus in a sense pre-empted. By contrast, however, in the spontaneous condition, the preceding sentence fragments proved unintelligible, and subjects demonstrated by their responses that they either found it impossible to use them or were led to pursue erroneous hypotheses on the basis of the context they perceived. In this condition, therefore, they initially accessed a very large set of candidates, especially in the Syll-2 condition. The value of stressed syllables in the face of such a large number of candidates activated to a low level is likely to be great.

In a recent paper, Mehta and Cutler (1988) suggested that although experiments using read and spontaneous speech appear to produce inconsistent results, the same mechanisms can be postulated to account for them. All that differs between the two speech modes, they argue, is the opportunities which arise for employing particular processing strategies. Indeed, considerations of parsimony suggest that this must be the case. The mechanisms which have been proposed to account for the experimental results presented in this study are consistent with Mehta and Cutler's suggestion. Various factors may cause fluctuations in the relative informativeness of speech signals. Between speech styles, large differences of intelligibility may be found; and even within a single speech style, the utility of the segments in the speech signal may not be uniform. Yet it is possible to account for apparently disparate behaviour using a single mechanism.

Chapter Notes

(1) These points are arbitrary with respect to linguistic categories.

(2) This is not to suggest that the presence of false starts, hesitations etc does not present serious problems for the human speech processing mechanism. For the purposes of this discussion, however, mechanisms for dealing with such phenomena will be treated as aberrations from normal processing behaviour, and ignored.

(3) One might ask, then, why the problem of automatic recognition of words in connected speech has not yet been solved even when the input is such as would gladden the heart of an elocution teacher. Note, however, that the state of the art in automatic linguistic and phonetic processing has not yet equalled that of the human word recognition device. S. Isard (personal communication) has pointed out that the only automatic devices which are currently achieving any success in word recognition (those using dynamic time warping and hidden markov modelling) are somewhat similar to the mechanism outlined above; these automatic mechanisms achieve this success despite apparently rather poor phoneme recognition scores.

(4) Although it accounts satisfactorily for much of the evidence discussed in this study, McClelland and Elman's (1986) TRACE model has almost nothing to say about the use of contextual information in the activation process. A plausible account might be that syntactic and semantic features derived from the preceding discourse are allowed to

activate candidates in a manner similar to that proposed for the acoustic-phonetic features. It would probably be necessary to assign a lower weight to the contributions of the former than to those of the latter, since acoustic-phonetic cues seem to constrain access more efficiently than top-down cues (Lowe 1988).

References

Abercrombie, D. 1967. *Elements of General Phonetics.* Edinburgh: Edinburgh University Press.

Adams, C. and Munro, R.R. 1977. In search of the acoustic correlates of stress; fundamental frequency, amplitude and and duration in the connected speech of native and non-native speakers. *Phonetica* 35, 125 - 156.

Altmann, G. and Carter, D. (in press). Lexical stress and lexical discriminability: stressed syllables are more informative, but why? *Computer Speech and Language.*

Bagley, W.C. 1900. The apperception of the spoken sentence: a study in the psychology of language. *American Journal of Psychology* 12, 80 - 130.

Balota, D. and Chumbley, J.I. 1984. Are lexical decisions a good measure of lexical access? The role of word frequency in the neglected decision stage. *Journal of Experimental Psychology: Human Perception and Processing* 10, 340 - 357.

Bansal, R.K. 1966. *The intelligibility of Indian English.* Unpublished PhD thesis, London University.

Bard, E.G. and Anderson, A.H. 1982. The unintelligibility of speech addressed to children. *Journal of Child Language* 10, 265 - 292.

Bard, E.G. (forthcoming). Competition, lateral inhibition and frequency. In G. Altmann (ed).

Bard, E.G., Shillcock, R.C. and Altmann, G.T.M. 1988. The recognition of words after their acoustic offsets: effects of subsequent context. *Perception and Psychophysics* 44 (5), 395 - 408.

Bloomfield, L. 1933. *Language.* New York: Holt.

Blutner, R. and Sommer, R. 1988. Sentence processing and sentence accent: the influence of the focus-identifying task. *Journal of Memory and Language* 27, 359 - 367.

Bolinger, D.L. 1958. A theory of pitch accent in English. *Word* 14, 109 - 149.

Bond, Z.S. 1971. Units in speech perception. *Ohio State University Working Papers in Linguistics* 9, 1 - 112.

Bond, Z.S. 1981. Listening to elliptic speech: pay attention to stressed vowels. *Journal of Phonetics* 9, 89 - 96.

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

Bond, Z. and Garnes, S. 1980. Misperceptions of fluent speech. In R. Cole (ed), *Perception and Production of Fluent Speech.* Hillsdale: Erlbaum.

Bridle, J., Brown, M.D. and Chamberlain, . 1982. A one-pass algorithm for connected speech recognition. *Proceedings of IEEE ICASSP*, 899 - 902.

Broadbent, D.E. 1967. Word-frequency effect and response bias. *Psychological Review* 74, 504 - 506.

Browman, C.P. 1978. Tip of the tongue and slip of the ear: implications for language processing. *UCLA Working Papers in Phonetics* 42.

Brown, G. 1977. *Listening to Spoken English.* London: Longman.

Brown, G. and Yule, G. 1983. *Discourse Analysis.* Cambridge: Cambridge University Press.

Brown, R. and McNeill, D. 1966. The tip-of-the-tongue phenomenon. *Journal of Verbal Learning and Verbal Behaviour* 5, 325 - 337.

Bruner, J. and O'Dowd, D. 1958. A note on the informativeness of words. *Language and Speech* 1, 98 - 101.

Butterfield, S. and Cutler, A. 1988. Segmentation errors by human listeners: evidence for a prosodic segmentation strategy. *Proceedings of SPEECH '88* 7th FASE Symposium, Edinburgh.

Calfee, R.C., Chapman, R.S. and Venezky, R.L. 1972. How a child needs to learn to think to learn to read. In L. W. Gregg (ed) *Cognition in Learning and Memory.* New York: Wiley.

Carter, D. 1987. An information-theoretic analysis of phonetic dictionary access. *Computer Speech and Language* 2, 1 - 11.

Cherry, E.C. 1953. Some experiments on the recognition of speech, with one and with two ears. *Journal of the Avoustical Society of America* 25, 975 - 979.

Chomsky, N. and Halle, M. 1968. *The Sound Pattern of English*. New York: Harper and Row.

Clark, H. 1973. The language-as-fixed-effect fallacy: a critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behaviour* 12, 335-359.

Cohen, J. and Cohen, P. 1983. *Applied Multiple Regression and Correlation Analysis for the Behavioral Sciences*. Hillsdale, N.J.: Erlbaum.

Cole, R. 1973. Listening for mispronunciations: a measure of what we hear during speech. *Perception and Psychophysics* 11, 153 - 156.

Cole, R. and Jakimik, J. 1978. Understanding speech: how words are heard. In G. Underwood (ed), *Strategies of Information Processing*. London: Academic Press.

Cole, R. and Jakimik, J. 1980(a). A model of speech perception. In R. Cole (ed), *Perception and Production of Fluent Speech*. Hillsdale: Erlbaum.

Cole, R. and Jakimik, J. 1980(b). How are syllables used to recognise words? *Journal of the Acoustical Society of America* 67, 965 - 970.

Cole, R., Jakimik, J. and Cooper, W.E. 1978. Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America* 64 (1) 44 - 56.

Cole, R. and Perfetti, C. 1980. Listening for mispronunciations in a child's story: the use of context by children and adults. *Journal of Verbal Learning and Verbal Behaviour* 19, 297 - 315.

Coltheart, M. 1981. The MRC psycholinguistic database. *Quarterly Journal of Experimental Psychology* 33a, 497 - 505.

Connine, C. 1987. Constraints on interactive processes in auditory word recognition: the role of sentence context. *Journal of Memory and Language* 26, 527 - 538.

Connine, C., Clifton C.E. and Cutler, A. 1987. Effects of lexical stress on phonetic categorisation. *Phonetica* 44, 133 - 146.

Cooper, W.E. and Paccia-Cooper, J. 1980. *Syntax and Speech.* Cambridge, Mass.: Harvard University Press.

Cooper, W.E. , Soares, C. and Ham, A. 1983. The influence of inter- and intra-speaker tempo differences on fundamental frequency and palatalisation. *Journal of the Acoustical Society of America* 73 (5) 1723 - 1730.

Cotton, S. and Grosjean, F. 1984. The gating paradigm: a comparison of successive and individual presentation formats. *Perception and Psychophysics* 35, 41 - 48.

Cox, S.J. 1988. Hidden Markov models for automatic speech recognition: theory and application. *British Telecom Technology Journal* 6 (2), 105 - 115.

Crain, S. and Steedman, M. 1985. On not being led up the garden path: the use of context by the sentence processor. In Dowty, D.R., Kartunnen, L. and Zwicky, A.M. (eds) *Natural language parsing -- psychological, computational and theoretical perspectives* Cambridge: CUP.

Crystal, D. and Davy, D. 1969. *Investigating English Style.* London: Longman.

Cutler, A. 1976. Phoneme monitoring reaction time as a function of preceding intontion contour. *Perception and Psychophysics* 20, 55 - 60.

Cutler, A. 1982. Prosody and sentence perception in English. In J. Mehler, E.C.T. Walker and M. Garrett (eds) *Perspectives on Mental Representation: Experimental and Theoretical Studies of Cognitive Processes and Capacities.* Hillsdale, N.J.: Erlbaum.

Cutler, A. 1986. *Forbear* is a homophone: lexical prosody does not constrain lexical access. *Language and Speech* 29, 201 - 220.

Cutler, A. and Carter, D. 1987. The predominance of strong initial syllables in English. *Computer Speech and Language* 2, 133 - 142.

Cutler, A. and Clifton, C.E. 1984. The use of prosodic information in word recognition. In H. Bouma and D.G. Bouwhuis (eds) *Attention and Performance* vol. 10. Hillsdale: Erlbaum.

Cutler, A. and Fodor, J.A. 1979. Semantic focus and sentence comprehension. *Cognition* 7, 49 - 59.

Cutler A. and Foss, J. 1977. On the role of sentence stress in sentence processing. *Language and Speech* 20, 1 - 10.

Cutler, A., Hawkins, J.A. and Gilligan, G. 1987. The suffixing preference: a processing explanation. *Linguistics* 23, 723 - 758.

Cutler, A. and Ladd, D.R. 1983. *Prosody: Models and Measurements* Berlin: Springer-Verlag.

Cutler, A. and Norris, D. 1988. The role of strong syllables in speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance* 14, 113 - 121.

Dalby, J. 1986. *Phonetic structure of fast speech.* Bloomington: Indiana Linguistics Club.

Davidoff, J.B., Beaton, A.A., Done, D.J.and Booth, H. 1982. Information extraction from brief visual displays: half-field and serial position effects in children, normal and illiterate adults. *British Journal of Psychology* 73, 29 - 39.

Dixon, W.J., Brown, M.B., Engelman, L., Frane, J.W., Hill, M.A., Jennrich, R.I. and Toporek, J.D. 1983. *BMDP Statistical Software.* Berkeley; University of California Press.

Ellis, A. and Beattie, G. 1986. *The Psychology of Language and Communication.* London: Weidenfeld and Nicolson.

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

Forster, K.I. 1976. Accessing the mental lexicon. In R.J. Wales and E.C.T.Walker (eds) *New Approaches to Language Mechanisms.* Amsterdam: North-Holland.

Forster, K.I. 1979. Levels of Processing and the structure of the language processor. In W.E. Cooper and E.C.T. Walker (eds) *Sentence Processing: Psycholinguistic studies in honour of Merrill Garrett.* Hillsdale, N.J.: Lawrence Erlbaum.

Forster, K.I. and Chambers, S.M. 1973. Lexical access and naming time. *Journal of Verbal Learning and Verbal Behaviour* 8, 457 - 462.

Foss, D. 1969. Decision processes during sentence comprehension: effects of lexical item difficulty and position upon decision times. *Journal of Verbal Learning and Verbal Behaviour* 8, 457 - 462.

Foss, D.J., Harwood, D.A. and Blank, M.A. 1980. Deciphering decoding decisions: data and devices. In R.A. Cole (ed) *Perception and Production of Fluent Speech.* Hillsdale, N.J.: Erlbaum.

Fowler, C.A. and Housum, J. 1987. Talkers' signaling of 'new' and 'old' words in speech and listeners' perception and use of the distinction. *Journal of Memory and Language* 26, 489 - 504.

Francis, W.N. and Kucera, H. 1982. *Frequency Analysis of English Usage: Lexicon and Grammar* Boston: Houghton Mifflin.

Fry, D. B. 1958. Experiments in the perception of stress. *Language and Speech* 1, 126 - 152.

Fry, D. B. 1979. *The Physics of Speech*. Cambridge: CUP.

Fudge, E. 1984. *English Word Stress*. London: George Allen and Unwin.

Gee, J. and Grosjean, F. 1983. Performance structures: a psycholinguistic and linguistic appraisal. *Cognitive Psychology* 15, 411 - 458.

Gimson, A.C. 1980. *An Introduction to the Pronunciation of English* (second edition) London: Edward Arnold.

Grosjean, F. 1980. Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 28, 267 - 283.

Grosjean, F. 1983. How long is the sentence? Prediction and prosody in the on-line processing of language. *Linguistics* 21 (3) 501 - 529.

Grosjean, F. 1985. The recognition of words after their acoustic offsets: evidence and implications. *Perception and Psychophysics* 38, 299 - 310.

Grosjean, F. and Gee, J. 1987. Prosodic structure and word recognition. *Cognition* 25, 135 - 155.

Grosjean, F. and Itzler, J. 1984. Can semantic constraint reduce the role of word frequency during spoken-word recognition? *Bulletin of the Psychonomic Society* 22(3), 180-182.

Grosjean, F. and Lane, H. 1981. Temporal variables in the perception and production of Spoken and Sign Languages. In P. Eimas and J. Miller (eds) *Perspectives on the Study of Speech* Hillsdale: Lawrence Erlbaum Associates.

Harrington, J., Watson, G. and Cooper, M. 1988. Word-boundary segmentation from phoneme sequence constraints in automatic continuous speech recognition. In D. Vargha (ed.) *Proceedings of the 12th International Conference on Computational Linguistics (Coling '88)* Volume 1, 225 - 229. John von Neumann Society for Computing Sciences: Budapest.

Hawkins, P. 1984. *Introducing Phonology* London: Hutchinson University Library.

Hayes, B. 1980. *A metrical theory of stress rules.* Unpublished PhD dissertation, MIT.

Hayes, B. 1982. Extrametricality and English stress. *Linguistic Inquiry* 13, 227 - 276.

van Heuven, V. 1988. Effects of stress and accent on the human recognition of word fragments in spoken context: gating and shadowing. *Proceedings of the Institute of Acoustics*, FASE Conference, Edinburgh.

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

Hunnicutt, S. 1985. Intelligibility versus redundancy -- conditions of dependency. *Language and Speech 28, 47 - 56.*

Huttenlocher, D.P. 1985. *Exploiting sequential phonotactic constraints in recognising spoken words.* Massachusetts Institute of Technology, Artificial Intelligence Laboratory; AI Memo 867.

Huttenlocher, D.P. and Zue, V.W. 1983. *Phonotactic and Lexical Constraints in Speech Recognition* Paper presented at the American Association for Artificial Intelligence, August 1983.

Hyman, L. 1977. *Studies in Stress and Accent.* Los Angeles: University of Southern California Linguisics Department.

Idrissi, B. 1987. *Metalinguistic awareness in literate and illitrate children and adults.* Unpublished doctoral thesis, University of Edinburgh.

Jakimik, J. 1979. *The interaction of sound and knowledge in word recognition from fluent speech.* Unpublished doctoral dissertation, Carnegie-Mellon University.

Johns-Lewis, C. 1986. Prosodic differentiation of discourse modes. In C. Johns-Lewis (ed.) *Intonation and discourse.* London: Croom Helm.

Johns-Lewis, C. 1987. The perception of discourse modes. In M. Coulthard (ed.) *Discussing discourse* University of Birmingham: Discourse Analysis Research Monograph 14, 249 - 271.

Kingdon, R. 1958. *The Groundwork of English Stress.* London: Longman.

Kiparsky, P. 1979. Metrical structure assignment is cyclic. *Linguistic Inquiry* 10, 421 - 422.

Kucera, H. and Francis, W.N. 1967. *Computational Analysis of Present-Day American English.* Providence: Brown University Press.

Ladefoged, P. 1982. *A Course in Phonetics.* Second edition. New York: Harcourt Brace Jovanovich.

Lagerquist, L.M. 1980. Linguistic evidence from paranomasia. *Papers from the Sixteenth Regional Meeting, Chicago Linguistic Society.*

Lamel, L. and Zue, V. 1984. *Properties of consonant sequences within words and across word boundaries.* Paper presented at IEEE ICASSP, San Diego, California.

Laver, J. 1980. *The phonetic description of voice quality.* Cambridge; Cambridge University Press.

Lehiste, I. 1960. An acoustic-phonetic study of internal open juncture. *Phonetica* 5, (supplement), 1 - 54.

Lehiste, I. 1967. *Suprasegmentals.* Cambridge, Mass.: MIT Press.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P. and Studdert-Kennedy, M. 1967. Perception of the speech code. *Psychological Review* 6, 431 - 461.

Liberman, I.Y., Shankweiler, D., Liberman, A.M., Fowler, C. and Fischler, F.W. 1977. Explicit phoneme and syllable segmentation in the young child. *Journal of Experimental Child Psychology, 18, 201-212.*

Liberman, M. and Prince, A. 1977. On stress and linguistic rhythm. *Linguistic Inquiry* 8, 249 - 336.

Lieberman, P. 1963. Some effects of semantic and grammatical context on the production and perception of speech. *Language and Speech* 6, 172 - 175.

Linell, P. 1982. *The Written Language Bias in Linguistics.* Studies in Communication, University of Linkoping.

Lowe, A. 1988. *The relative contributions of top-down and bottom-up information during lexical access.* Unpublished doctoral thesis, University of Edinburgh. Luce, P.A. 1986. A computational analysis of uniqueness points in auditory word recognition. *Prception and Psychophysics* 34, 155 - 158.

Marslen-Wilson, W. 1973. Linguistic structure and speech shadowing at very short latencies. *Nature* 244, 522 - 523.

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

Marslen-Wilson, W. 1980. Speech understanding as a natural process. In J. Simon (ed) *Spoken Language Understanding and Generation.* Dordecht: Reidel.

Marslen-Wilson, W. 1984. Function and process in spoken word recognition. In H. Bouma and D.G. Bouwhuis (eds) *Attention and Performance 10: Control of Language Processes*

Marslen-Wilson, W. 1987. Functional parallelism in spoken word recognition. *Cognition* 25, 71 - 102.

Marslen-Wilson, W. 1988. Activation, competition and frequency in lexical access. Paper presented at the Experimental Psychology Society, Edinburgh, July.

Marslen-Wilson, W., Brown, C.M. and Zwitserlood, P. (in preparation). Spoken word-recognition: early activation of multiple semantic codes.

Marslen-Wilson, W. and Tyler, L.K. 1980. The temporal structure of spoken language understanding. *Cognition* 8, 1 - 71.

Marslen-Wilson, W., Tyler, L.K. and Seidenberg, M. 1978. The semantic control of sentence segmentation. In W.J.M. Levelt and G.B. Flores d'Arcais (eds) *Studies in the Perception of Language*. London: Wiley. Hillsdale, NJ: Erlbaum.

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

McCarthy, J. 1979. On stress and syllabification. *Linguistic Inquiry* 10, 433 - 466.

McClelland, J.L. and Elman, J.L. 1986. Interactive processes in speech perception: the TRACE model. In D. Rumelhart and J.L. McClelland (eds) *Parallel Distributed Processes: Explorations in the Microstructure of Cognition.* Cambridge, Mass.: MIT Press.

McClelland, J.L. and Rumelhart, D.E. (1988) *Explorations in Parallel Distributed Processing: a Handbook of Models, Programs and Exercises.* Cambridge, MA: MIT.

Mehta, G. and Cutler, A. 1988. Detection of target phonemes in spontaneous and read speech. *Language and Speech* 31 (2), 135 - 156.

Mehler, J., Segui, J. and Carey, P. 1978. Tails of words: monitoring ambiguity. *Journal of Verbal Learning and Verbal Behaviour* 17, 29 - 35.

Miller, G.A., Heise, G.A. and Lichten, W. 1951. The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology* 41, 329 - 335.

Miller, G.A. and Isard, S. 1963. Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behaviour* 2, 217 - 228.

Miller, G.A. and Nicely, P. 1955. An analysis of perceptual confusions among some English consonants. *Journal of the Acoustical Society of America* 27, 338 - 352.

Mohanan, K.P. 1986. *The Theory of Lexical Phonology*. Dordecht: Reidel.

Morais, J., Cary, L., Alegria, J. and Bertelson, P. 1979. Does awareness of speech as a sequence of phones arise spontaneously. *Cognition* 7, 323 - 331.

Moray, N. 1959. Attention in dichotic listening: affective cues and the influence of instructions. *Quarterly Journal of Experimental Psychology* 11, 56 - 60.

Morton, J. 1969. Interaction of information in word recognition. *Psychological Review* 76, 165 - 178.

Morton, J. 1979. Word Recognition. In J. Morton and J. Marshall (eds) *Psycholinguistcs Series 2: Structures and Processes* London: Eleck.

Morton, J. 1982. Disintegrating the lexicon: an information processing approach. In J. Mehler, E.C.T. Walker and M.F. Garrett (eds) *Perspectives on Mental Representation: Experimental Studies of Cognitive Processes and Capacities.* Hillsdale, N.J.: Erlbaum.

Morton, J. and Long, J. 1976. Effect of word transitional probability on phoneme identification. *Journal of Verbal Learning and Verbal Behaviour* 15, 43 - 51.

Nakatani, L.H. and Dukes, K.D. 1977. Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America* 62, 714 - 719.

Nolan, F. 1983. *The phonetic bases of speaker recognition* Cambridge; Cambridge University Press.

Nooteboom, S.D. 1981. Lexical retrieval from fragments of spoken words: beginnings versus endings. *Journal of Phonetics* 9, 407 - 424.

Norris, D. 1982. Autonomous processes in comprehension: a reply to Marslen-Wilson and Tyler. *Cognition* 11, 97 - 101.

Norris, D. and Cutler, A. 1985. Juncture detection. *Linguistics* 23, 689 - 705.

Pedlow, J. and Wales, R. 1987. Some effects of context on spoken language production. *Language and Speech* 30, 373 - 385.

Peterson, G.E. and Barney, H.L. 1952. Control methods used in a study of the vowels. *Journal of the Acoustical Society of America* 24, 175 - 184.

Peterson, G. E. and Lehiste, I. 1960. Duration of syllable nuclei in English. *Journal of the Acoustical Society of America* 32, 693 - 703.

Pickett, J.M. and Pollack, I. 1963. Intelligibility of excerpts from fluent speech: effects of rate of utterance and duration of excerpt. *Language and Speech* 6, 151 - 164.

Pisoni, D.B. 1981. Some theoretical issues in speech perception. *Cognition* 10, 249 - 259.

Pollack, I. and Pickett, J.M. 1963. The intelligibility of excerpts from conversation. *Language and Speech* 6, 165 - 171.

Rabiner, L.R. and Juang, B.H. 1986. An introduction to Hidden Markov Models. *EEE ASSP Magazine* January 1986.

Rubenstein, H., Garfield, L. and Millikan, J.A. 1970. Homographic entries in the internal lexicon. *Journal of Verbal Learning and Verbal Behaviour* 9, 487 - 494.

Salasoo, A. and Pisoni, D.B. 1985. Interaction of knowledge sources in spoken word identification. *Journal of Memory and Language* 24 (2), 210 - 231.

Savin, H.B. and Bever, T.G. 1970. The nonperceptual reality of the phoneme. *Journal of Verbal Learning and Behaviour* 9, 295 - 302.

Schubiger, Maria. 1967. Vowel quality in unstressed syllables. *English Language Teaching* 21, 2.

Scott, D. and Cutler, A. 1984. Segmental phonology and the perception of syntactic structure. *Journal of Verbal Learning and Verbal Behaviour* 23, 450 - 466.

Selkirk, E. 1978. On prosodic structure and its relation to syntactic structure. In T. Fretheim (ed) *Nordic Prosody II*. Trondheim: TAPIR.

Selkirk, E. 1980. The role of prosodic categories in English word stress. *Linguistic Inquiry* 11, 563 - 605.

Shields, J., McHugh, A. and Martin, J. 1974. Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology* 102, 250 - 255.

Shillcock, R. (in press). Lexical hypotheses in continuous speech recognition. In G.T. Altmann (ed) *Cognitive models of speech processing: psycholinguistic and computational perspectives*. Cambridge, Massachussetts: MIT Press.

Shillcock, R.C., Bard, E.G. and Spensley, F. 1988. Some Prosodic effects on human word recognition in continuous speech. *Proceedings of the Institute of Acoustics FASE Conference*, Edinburgh.

Shipman D.W. and Zue, V. 1982. Properties of Large Lexicons: Implications for Advanced Isolated Word Recognition.
Proceedings of ICASSP, 546-549. Paris, France.

Shockey, L. 1983. *Phonetic and phonological properties of connected speech.* Ohio State Working Papers in Linguistics.

Shockey, L. Forthcoming. *Conversational Speech Phonology.*

Shockey, L. and Bond, Z.S. 1980. Phonological processes in speech addressed to children. *Phonetica* 37, 267 - 274.

Simon, H. 1969.*The Sciences of the Artificial.* Cambridge, Massachussetts; MIT Press.

Signal Technology Inc. 1985. *ILS User's guide V5.0.*

Stampe, D. 1979. *A dissertation on Natural Phonology.* New York: Garland.

Stanners, R.P., Neisser, J.J., Hernon, W.P. and Hall, R. 1979. Memory representation for morphologically relate words. *Journal of Verbal Learning and Verbal Behaviour* 18, 399 - 412.

Svartvik, J. and Quirk, R. 1980. *A Corpus of English Conversation.* London: Gleerup.

Swinney, D. 1979. Lexical acccess during sentence comprehension: (re)consideration of context effects. *Journal of Verbal Learning and Verbal Behaviour* 18, 645 - 659.

Taft, L. 1984. *Prosodic constraints and lexical parsing strategies.* Unpublished doctoral dissertation, University of Massachussetts.

Terry, M., Hiller, S., Laver, J. and Duncan, G. 1986. The AUDLAB interactive speech analysis system. *Proceedings of IEEE International Conference on Speech I/O: Techniques and Applications.*

Tyler, L.K. 1984 . The structure of the word-initial cohort: evidence from gating. *Perception and Psychophysics* 36, 417 - 427.

Tyler, L.K. and Marslen-Wilson, W. 1986. The effects of context on the recognition of polymorphemic words. *Journal of Memory and Language* 25, 741 - 752.

Tyler, L.K. and Warren, P. 1987. Local and global structure in spoken language comprehension. *Journal of Memory and Language* 26, 638 - 657.

Tyler, L.K. and Wessels, J. 1983. Quantifying contextual contributions to word-recognition processes. *Perception and Psychophysics* 34, 409 - 420.

Tyler, L.K. and Wessels, J. 1985. Is gating an on-line task? *Perception and Psychophysics* 38, 217 - 222.

Umeda, N. 1977. Consonant durations in American English. *Journal of the Acoustical Society of America* 61, 846 - 858.

Vaissiere, J. 1985. Speech recognition: a tutorial. In F. Fallside and W. Woods (eds) *Computer Speech Processing.* London: Wiley.

Wells, J.C. 1982. *Accents of English* Cambridge: Cambridge University Press.

Wheeldon, L. 1985. *The intelligibility of words from three different kinds of speech.* Unpublished undergraduate disseration, University of Edinburgh Department of Linguistics.

Williams, C.E. and Stevens, K.N. 1972. Emotions and speech: some acoustical correlates. *Journal of the Acoustical Society of America* 52, 1238 - 1250.

Appendix A: Frequencies* of Test Words, Experiment 1.

Stressed on the first syllable		Stressed on the second syllable	
substitute	17	subscription	4
musical	88	musician	65
temperature	161	temptation	18
discipline	29	discovery	55
dictionary	59	distinction	56
property	222	professor	78
pyramid	1	pyjamas	4
caravan	5	cathedral	11
communist	112	committee	188
spectacle	21	spectator	22
carpenter	11	cartoonist	4
supplement	21	supporter	11
dynamite	6	diversion	11
subsidy	7	suggestion	57
surgery	6	survivors	15
cemetery	15	sensation	24
vacancy	8	vacation	4
compliment	3	compartment	12
parliament	17	partition	7
company	453	computer †	18
tragedy	56	translation	19
masterpiece	12	magician	7
banister	7	banana	5
comedy	41	commission	117
	Mean = 57.4		Mean = 33.8
	S.D. = 100.6		S.D. = 43.9

* Francis and Kucera 1982

† It seems unlikely that this item would yield such a low score in a frequency count conducted on present-day materials. Excluding this score and its partner from the analysis yields the following descriptive statistics: Syll-1 mean 40.22, S.D. 56.27; Syll2 mean 34.52, S.D. 44.72.

Appendix B: Stimuli, Experiment 1

Notes:

(i) Test words are italicised: words following test words were not presented to subjects (see text).

(ii) Sentences are presented in pairs for ease of comparison of related items: this reflects neither the order in which items were recorded nor the order of presentation in the experiment.

1(a) At half time, the coach sent for a *substitute* in a hurry.

1(b) When she joined the sports club, Sue had to pay a *subscription* on the first day.

2(a) Last time he went to the theatre, John went to see a *musical* about cats.

2(b) After practising the piano for many years, Paul became a *musician* in an orchestra.

3(a) Because the room seemed so hot, John checked the *temperature* on the thermostat.

3(b) For someone on a diet, a cream cake is a *temptation* at any time.

4(a) Although he wanted to join the army, Mark couldn't take the *discipline* of the training.

4(b) Working alone in his laboratory, the scientist made the *discovery* with no assistance.

5(a) When she started her French course, Ann bought a *dictionary* from the bookshop.

5(b) Although beer and lager are quite similar, you should make a *distinction* between them.

6(a) When house prices were high, the agent sold the *property* for fifty thousand.

6(b) On my first day at university, I met the *professor* for the first time.

7(a) In the desert, the king built the *pyramid* for his burial.

7(b) In his bedroom, he put on the *pyjamas* with the yellow stripes.

8(a) On holiday in Wales, the kids stayed in the *caravan* for a week.

8(b) On my trip to Canterbury, I visited the *cathedral* for a while.

9(a) After stealing the plans, the spy sold them to a *communist* for a large sum.

9(b) To look into the problem, Margaret set up a *committee* of MPs.

10(a) When Charles and Diana got married, it was quite a *spectacle* for all to see.

10(b) When John threw his racquet in the air, it nearly hit a *spectator* on the head.

11(a) Because he liked making things, Tom became a *carpenter* after leaving school.

11(b) Because she liked drawing things, Ann became a *cartoonist* after leaving school.

12(a) With our Sunday paper, we get a *supplement* in lurid colour.

12(b) Although he sometimes watches our team, Bill is not a *supporter* in any sense.

13(a) To demolish the chimney, we had to use the *dynamite* from our supplies.

13(b) Because of the roadworks, we had to take the *diversion* for many miles.

14(a) To encourage them to employ young people, firms can get a *subsidy* from the government.

14(b) To help the kids solve the problem, the teacher made a *suggestion* about it.

15(a) When I went to see the doctor, he was in the *surgery* with a patient.

15(b) After the crash, we looked for the *survivors* in the wreckage.

16(a) Although she was afraid of ghosts, Jane walked through a *cemetery* at night.

16(b) When the story came out about Andrew, it caused quite a *sensation* in the press.

17(a) When the secretary left, we advertised a *vacancy* in the local paper.

17(b) When we won the money, we went on a *vacation* for a month.

18(a) In her new dress, Jane expected a *compliment* from her husband.

18(b) Boarding the train in London, Fred found a *compartment* at once.

19(a) When the country became independent, it had already had a *parliament* for many years.

19(b) To divide up the open-plan office, the joiner made a *partition* out of wood.

20(a) With his redundancy money, Jim formed a *company* of his own.

20(b) To send out bills, the firm uses a *computer* for economy.

21(a) Because he was so young, Jim's accident was a *tragedy* for his parents.

21(b) Because the jury couldn't understand French, they needed a *translation* of the evidence.

22(a) When I found the old painting, I knew it was a *masterpiece* at once.

22(b) When he shows us his tricks, we all admire a *magician* for his skill.

23(a) As she came downstairs, she held the *banister* with her hand.

23(b) From all the fruit in the bowl, John chose the *banana* for his lunch.

24(a) On television last night, we watched a *comedy* about doctors.

24(b) When he sold the double glazing, Jim earned a *commission* for his work.

Appendix C: Practice items, Experiment 1 (gated word in italics).

1. (Heard in Long Context form) Before building the new factory, we had to choose a *location* for it.

2. (Heard in Short Context form) For the high unemployment figures, everyone blames the *government* because of its policy.

3. (Heard in No Context form) For her letters, she uses a *typewriter* in her office.

Appendix D: contextual appropriateness of responses at gate 1, Experiment 1.

Note: * $p < .05$
** $p < .01$

Table D1: Percentage of responses at gate 1 of Experiment 1 which were nouns

STRESS:	Syll-1			Syll-2		
CONTEXT:	none	short	long	none	short	long
LITERACY:						
High	79	88	94	69	85	90
Low	77	80	85	65	78	81
Mean	78	84	90	67	82	86

Table D2: ANOVA for data in Table D1

Source	F1	F2	Min F′
Literacy	(1,34) 5.73 *	(1,46) 11.44 **	(1,65) 3.82 n.s.
Stress	(1,34) 9.72 **	n.s.	---
Context	(1,34) 19.10 **	(1,46) 10.24 **	(1,156) 6.67 *

Context means: all differences significant at $p < .01$, by-subjects Scheffé test

Table D3: Percentage of responses in No Context and Short Context conditions which were judged appropriate completions of the Short Context sentence (gate 1, Experiment 1)

STRESS:	Syll-1		Syll-2	
CONTEXT:	none	short	none	short
LITERACY:				
High	28	75	15	69
Low	22	59	13	56
Mean	23	67	14	62

Table D4: ANOVA for data in Table D3

Source	F1	F2	Min F′
Literacy	(1,34) 12.66 **	(1,46) 8.16 **	(1,80) 4.96 *
Stress	(1,34) 8.83 **	n.s.	---
Context	(1,34) 385.12 **	(1,46) 110.76 **	(1,69) 86.02 **
Context X Literacy	(1,34) 5.83 *	n.s.	---

Table D5: Percentage of responses in No Context and Long Context conditions which were judged appropriate completions of the Short Context sentence (gate 1, Experiment 1)

STRESS:	**Syll-1**		**Syll-2**	
CONTEXT:	**none**	**long**	**none**	**long**
LITERACY:				
High	6	75	2	81
Low	3	58	2	64
Mean	5	67	2	73

Table D6: ANOVA for data in Table D5

Source	**F1**	**F2**	**Min F′**
Literacy	(1,34) 27.28 **	(1,46) 16.77 **	(1,79) 10.39 **
Stress	n.s.	n.s.	---
Context	(1,34) 962.08 **	(1,46) 374.17 **	(1,74) 269.40 **
Context X Literacy	(1,34) 13.42 **	(1,46) 12.71 **	(1,79) 6.53 *

Context X Literacy interaction: the No Context means differed at $p < .05$, while all other comparisons reached significance at $p < .001$ (by-subject Scheffé test).

Appendix E: Testwords, Experiment 2

1. Speaker 1

SYLL1.	SYLL2.
opening	exhausted
quantities	temptation
anyway	today
leather	o'clock
nursery	upset
details	explain
message	suppose
parcel	collected
fabulous	impediment
naughty	delicious
cousin	mobility
strangers	tonight
photographs	amazing
always	routine
everything	reported
mother	believe
funeral	eventually
Tuesday	amount
punishment	exact
contract	impression
little	enough
money	decide
Friday	abroad
accident	confirm

Speaker 2

SYLL1.	SYLL2.
heavily	nearby
hospital	referring
jumpers	monopolised
waitressing	immaculate
pirates	again
bottom	imagine
noticed	remember
seventy	abrupt
forty	arrange
everyone	outside
probably	important
actually	relate
absolutely	asleep
lovely	inquiries
husband	surprised
noisy	embarassed
massive	divorce
coffin	procession
easily	capacities
ready	downhill
seven	relationship

definitely	tomorrow
hangover	unusual
merry	eleventh

Speaker 3

SYLL1.	SYLL2.
manager	arrived
April	command
difficult	especially
typical	ignore
awful	presume
radio	direct
finally	event
travel	afraid
scenery	invisible
organising	address
signatures	excited
weather	ideal
shotguns	disgusting
terrible	intelligence
railway	arena
obviously	affair
cancelled	originally
quality	account
happens	society
mentioned	avoided
nephew	unpopular
various	away
lunchtime	hilarious
paper	support

Speaker 4

SYLL1.	SYLL2.
Christmas	museum
family	included
wedding	reminds
distance	hotel
damaged	connecting
pavement	deposit
brochure	brigade
realised	police
candles	became
village	engaged
beautifully	incentive
narrow	delivered
teenagers	discovered
charities	departure
holiday	produced
careful	historical
furniture	piano
jewellery	October
camera	complain

ordinary	refused
title	admit
single	particular
certainly	reduced
advertised	guitar

Speaker 5

SYLL1	SYLL2.
rotten	prevents
salary	familiar
formal	succeed
language	enjoy
comic	hysterical
sentences	extremely
educated	cartoon
interest	control
common	machine
building	applying
plastic	acceptance
logical	alone
social	concerned
foreign	production
beautiful	research
interview	convince
pregnancy	addictive
handy	react
medicine	upstairs
public	developed
makeup	linguistics
summer	entire
accent	along
normally	publicity

Speaker 6

SYLL1.	SYLL2.
mesmerised	mundane
dictionaries	appointed
knowledge	discussed
technical	sustain
challenge	afford
concentrate	pretend
settle	exam
double	apart
afterwards	impress
panic	inherited
English	degree
honours	resist
threshold	offended
coffee	description
comments	forgotten
valuable	persuade

moment	**consider**
utterly	**excuse**
alcohol	**phonology**
festival	**suspicion**
headache	**offensive**
several	**already**
modern	**revealed**
anecdote	**awake**

Appendix F: Orientation sentences, Experiments 2 and 3

Speaker 1

But she said if Dorothy puts her in the car and takes her down there she said she doesn't mind it but you know it's just sort of the physical thing of getting herself there.

Speaker 2

He has a y... a kitkat, fruit, crisps, and the sandwiches and something else doesn't he?

Speaker 3

You know you cannae make up your mind if they're pulling it down or putting it up.

Speaker 4

Some mornings we used to go out early and there was a lovely big, we would call it a loch, but I suppose they would call it a pond.

Speaker 5

I feel really sorry for men at times, I mean you can get to know women so quickly, you just take it for granted you've got so much in common.

Speaker 6

I haven't read any of it - the only thing I knew was on Spitting Image.

Appendix G: Percentage of trials resulting in identification of test word, Experiment 2 (inflectional suffixes ignored)

Table H1: Cell means, Experiment 2

VERSION:	Read			Spontaneous		
STRESS:	Syll-1	Syll-2	Mean	Syll-1	Syll-2	Mean
SPEAKER:						
1	86	83	(85)	81	45	(63)
2	78	68	(73)	54	52	(53)
3	75	73	(74)	46	30	(38)
4	86	71	(79)	71	66	(69)
5	90	82	(86)	53	40	(47)
6	84	67	(76)	66	47	(57)
Mean	83	74	(79)	62	47	(55)

Table H2: ANOVA, Experiment 2

Source	F1	F2	Min F′
Version	(1,23)135.97**	(1,276)102.66**	(1,180)58.50**
Stress	(1,23) 18.06**	(1,276) 15.60**	(1,96) 8.37**
Speaker	(5,115) 7.64**	(5,276) 2.91**	(5,390)2.11 n.s.
Vers. X Sp.	(5,115) 7.09**	(5,276) 3.44**	(5,389)2.32*
Vers. X Str. X Sp.	(5,115) 3.66**	n.s.	-----

**p<0.01 *p<0.05

Appendix H: Testwords (in upper case) and sentential contexts, Experiment 3

Speaker 1

Syll-1

They're OPENING on Mondays, which they don't usually.

Makes you think though doesn't it, all these lot coming in there with the LEATHER gear on.

Jane said something to me about that phone call, about someone from the other NURSERY coming up last Friday.

She wouldn't leave a MESSAGE, just 'oh, I'll phone another time'.

I got a big PARCEL from my Mum (on Tuesday.)

I had a FABULOUS choice in the end.

You've got PHOTOGRAPHS of yourself from that time.

Did we tell you about the FUNERAL the other day?

Syll-2

Andrew came over after you COLLECTED Martin.

Oh it's DELICIOUS, absolutely gorgeous.

It's AMAZING though isn't it.

We reckon they've been REPORTED to the health authorities.

Say you had the same AMOUNT of money as someone going to town.

I think it's brilliant how he does IMPRESSIONS of the kids in playgroup.

I'm sure she must buy a jumper with a hole in one sleeve and DECIDE to cut (the other one out to match it.)

Just to CONFIRM everything in writing.

Speaker 2
Syll-1

We can't drink HEAVILY because we can't have a hangover tomorrow morning.

Go to the BOTTOM of the hill.

I NOTICED she had a wedding ring and an engagement ring.

The man who died was SEVENTY seven.

It's PROBABLY just the same.

They didn't bring the COFFIN into (the) church.

They DEFINITELY needed us yesterday.

I said to him, and you're one of Robin Hood's MERRY men.

Syll-2

I REMEMBER we had this French test.

It was a bit ABRUPT, I felt.

She said I'll try to ARRANGE a lift.

So we got OUTSIDE and everyone was offering people lifts.

Trying to RELATE it to the (family tree.)

I think they're just going to make ENQUIRIES while they're over there.

I got EMBARASSED because they could all speak French.

Our CAPACITIES have gone down (haven't they?)

Speaker 3
Syll-1

You get an AWFUL lot of kids though.

I FINALLY got my typewriter fixed.

He started ORGANISING holidays.

But somebody said it's just like the Fife coast with good WEATHER and I think that's true.

All the people who had CANCELLED their milk for three days, they burgled all their houses.

Quite a lot of the stuff was good QUALITY stuff.

And something HAPPENS, you know.

So I'd to wrap it up in Christmas PAPER and he got it on Christmas morning.

Syll-2

This woman ARRIVED about half an hour later.

Obviously the person who had written it had a good COMMAND of the English language.

I PRESUME there were some.

It's frightening to think that your ADDRESS was being passed on to other people.

I was saying this is DISGUSTING, this is terrible.

I've joined him in the Building SOCIETY for his birthday.

You'll need to come and SUPPORT me when I'm at the Charity Fair in April.

She said they stopped outside the ARENA.

Speaker 4
Syll-1

It's their CHRISTMAS Fair but it's usually at the end of October in the Assembly rooms.

Swam all that DISTANCE and back again.

Which VILLAGE were you in?

The hotel's about ten minute's along a NARROW dirt track.

She saw FURNITURE being piled out and realised it was hers.

His fiancee gave him an ORDINARY camera for his Christmas.

His name got larger than the TITLE of the book.

She's still SINGLE, she's engaged now.

Syll-2

That REMINDS me, a couple of years ago I bought some Christmas cards.

There's some very expensive, big HOTELS in Crete.

The fire BRIGADE had come.

They DISCOVERED that the ones with the dots had dogs.

I like HISTORICAL fiction.

And then they'd COMPLAIN because there was maybe a tear (in the seam or something.)

I just REFUSED to touch it.

They were REDUCED from two pounds fifty to one twenty five.

Speaker 5
Syll-1

I used to get slagged ROTTEN because of my legs.

You should see some of the SENTENCES we got out of that.

I think we ought to have some kind of COMMON room for us.

If lunch wasn't such a SOCIAL thing with me I'd give it up.

She won't wear gloves on PUBLIC duty.

I often find I start wearing MAKEUP more in winter.

If I leave my lips with nothing even during the SUMMER then they start to crack.

You never think that there's any other reason for the ACCENT apart from sounding ridiculous.

Syll-2

Even if you don't know that much more you're FAMILIAR with it.

It's very important for him to SUCCEED isn't it?

I really ENJOY a good cry every now and then.

Any CARTOON's stereotyped.

It's sort of some CONTROL of bodily functions anyway.

As far as I'm CONCERNED they're something that just appears.

If I go to a new school I'll CONVINCE them all I'm totally different.

You can't say to the ENTIRE class let's meet here.

Speaker 6
Syll-1

I was just MESMERISED with fear.

They don't seem to like it if you actually ask them what they're talking about or CHALLENGE them at some point.

It always takes me at least six to eight months to actually SETTLE in somewhere.

I used to PANIC about it.

One passed in the exam with first class HONOURS and two failed and one dropped out.

At the MOMENT he stays with me two or three times a week.

I've got a HEADACHE dear it's from working.

Extensive knowledge of several MODERN languages or advanced training in phonetics.

Syll-2

I mean you can't get much more MUNDANE than that.

She was supposed to meet him somewhere at an APPOINTED time.

He said it's the only time he's ever DISCUSSED a relationship for as long as he's actually been going out with someone.

You can't SUSTAIN a relationship on seeing one another three nights a week and lunchtimes.

Quite apart from the fact that we can't f.. can't afford it he won't.

He's probably turned the speakers up so he can hear us anyway and PRETEND he's not.

He starts immediately to IMPRESS you with his amazing technical knowledge about computers.

And then he turned round afterwards and REVEALED this little anecdote about people speaking Welsh and English.

Appendix I: Published paper based on the work of this thesis

The role of context in auditory word recognition
J.M. McAllister
Perception and Psychophysics 1988, 44 (1), 94 - 97

Perception & Psychophysics
1988, 44 (1), 94-97

The use of context in auditory word recognition

J. M. McALLISTER
University of Edinburgh, Edinburgh, Scotland

This paper reexamines Tyler's (1984) hypothesis that hearers do not use contextual information during their processing of the early parts of auditorily presented words. In the gating experiment described here, identical word tokens were heard in a no-context condition and two with-context conditions. The data from the no-context condition provided a measure of the likelihood that a response with particular semantic and/or syntactic characteristics would occur by chance (i.e., without the influence of contextual factors). A comparison of these data with those produced in the with-context conditions revealed that hearers made use of contextual information even during the processing of the first 50 msec of test words. These results are discussed in relation to current theories of word recognition.

Models of auditory word recognition may be characterized according to the role played by contextual (syntactic, semantic, and pragmatic) factors. A crucial feature of the architecture of serial autonomous models (e.g., Forster, 1981) is the lack of reliance on top-down information in the word-recognition process; words are identified purely on the basis of their acoustic shape and are subsequently integrated into the developing syntactic and semantic representation of the sentence. Various pieces of experimental evidence (e.g., Marslen-Wilson & Tyler, 1980; Pollack & Pickett, 1963) have cast doubt on the validity of this type of model; it will not be considered further in this paper.

By contrast, interactive models of word recognition allow contextual information to contribute to the recognition process. For example, the logogen model (Morton, 1969) embodies an array of word-recognition elements, or logogens, that are sensitive to the presence of both acoustic-phonetic and contextual information that is consistent with a given word hypothesis; in the logogen model, information of either kind may be used as it becomes available. This is not the case for a second type of interactive theory, the cohort model (Marslen-Wilson, 1987; Marslen-Wilson & Tyler, 1980; Marslen-Wilson & Welsh, 1978). A defining feature of this model is the assembling of word candidates on the basis of an acoustic-phonetic match between their initial portion and that of the input;[1] this set of word candidates is termed the *word-initial cohort*. Subsequently, contextual and/or acoustic information is used to narrow down the word-initial cohort until a single candidate remains and recognition occurs. The cohort model therefore makes testable predictions about the role of contextual information in auditory word recognition. In particular, it should be possible to identify two phases: a first, bottom-up autonomous phase, in which context effects are absent, and a second, interactive phase, in which such effects can be detected.

Tyler (1984) used the gating paradigm to test some of the predictions of the cohort model. The gating paradigm, which was introduced by Grosjean (1980), is an experimental technique whereby a word fragment is presented repeatedly and is incremented by some prespecified amount at each successive pass. The subject's task is to identify the intended word. Thus, the first presentation (or gate) of a test word may consist of the first 50 msec of the word, the second gate of the first 100 msec, and so on until all of the word has been presented. On the assumption that the pooled responses of all subjects hearing a given word fragment are roughly equivalent to the set of words accessed by a single subject, researchers have used the gating paradigm to examine the pattern of search prior to the recognition of a word.

In Tyler's (1984) experiment, a single word-token was presented in five experimental conditions: a no-context condition, in which the word was heard in isolation, and four with-context conditions defined by systematically varying the semantic characteristics of the sentence (anomalous vs. minimal semantic constraints) and its syntactic characteristics (weak vs. strong syntactic constraints). Tyler estimated the length of the bottom-up autonomous phase at 150 msec; in her experiment, this corresponded to the first three gates of any stimulus. She analyzed several dependent variables, but those that are of interest for the purposes of the present discussion are those relating to the syntactic and semantic characteristics of the responses offered during these early gates. She found that, of the data collected from the first gate of each of her with-context conditions, on average 10% were syntactically inappropriate, and approximately 6% of the responses offered during the first three gates were semantically inappropriate.

Tyler (1984, p. 423) interpreted these results as support for the cohort model's claim that contextual factors are not allowed to influence cohort access, arguing that,

Address correspondence to J. M. McAllister, Centre for Speech Technology Research, 80 South Bridge, Edinburgh EH1 1HN, Scotland.

if contextual preselection were involved in the accessing of word candidates, *no* responses offered during these early gates would be semantically or syntactically inappropriate.

In fact, this argument is flawed on two counts, the first theoretical and the second methodological. First, the lack of temporal constraint on the use of top-down information need not imply that hearers will be able to make full use of such information on every occasion. For example, it will sometimes be the case that hearers will misinterpret contextual cues, and will access words consistent with their erroneous interpretation. Since the subjects' task in Tyler's (1984) experiment was only to identify the test word, no feedback was provided as to the subjects' understanding of the carrier sentence; therefore, it is possible that, at least in some instances, the subjects were responding on the basis of top-down information other than that present in the intended context. In such circumstances, it would not be surprising if subjects' responses were judged as inappropriate to the context. This issue is discussed in more detail below.

The second point concerns Tyler's (1984) method of analysis. Tyler excluded the no-context data from her contextual analyses, reasoning that only data produced in the with-context conditions were relevant to the assessment of the contributions of top-down factors. However, it is impossible to make claims regarding the use of context purely on the basis of such data, in the absence of an appropriate control condition. Without such a condition, the statistical significance of the percentages of contextually inappropriate responses cannot be evaluated.

In the experiment described in this paper, the cohort model's prediction that context effects will be absent during the processing of the initial portion of words was reexamined by comparing with-context data with data collected in a no-context control condition. That is, the responses offered in the no-context condition were compared with those offered in the with-context conditions, on the assumption that the no-context data would reflect the likelihood that syntactically or semantically appropriate candidates would occur by chance; any significant improvement over the no-context mean in the with-context conditions would therefore be attributable to the subjects' use of contextual information.

METHOD

Subjects

Thirty-six native speakers of Scots English volunteered to take part in the experiment. Each subject participated in a single experimental session lasting approximately 75 min, with a break approximately halfway through the session.

Materials

The test words were 48 trisyllabic nouns. Word frequencies ranged from 1 to 453 per million words of text (Francis & Kučera, 1982); the mean word frequency was 45.6. Contexts for the test words were constructed according to principles similar to those adopted by Grosjean (1980). Each test word was embedded in the second clause of a two-clause sentence whose structure was such that the second clause could stand alone as a well-formed sentence. An example of such a sentence (for the test word *surgery*) is as follows:

When I went to see the doctor, he was in his surgery with a patient.

The test word was preceded by between four and six syllables in the main clause and was followed by a prepositional or adverbial phrase. The sentences remained semantically and syntactically well-formed even after the removal of the final phrase.

Each sentence was used to generate three stimuli: a long-context sentence, consisting of the original sentence without the final phrase; a short-context sentence, derived from the long-context sentence by the deletion of the initial clause; and a no-context stimulus, which consisted of the test word alone.

The sentences were then pretested, using a visual cloze procedure, to determine the extent to which they predicted the test word. The long-context and short-context stimuli, up to and including the word before the test word, were presented to 20 subjects, who were asked to complete each sentence with a single word. Their responses were then compared with the intended word by an independent panel of three judges, and scored according to the metric suggested by Marslen-Wilson and Welsh (1978): 1 for a word identical to the target, 2 for a synonym, 3 for a related word, and 4 for a totally unrelated word. From these scores, two means were computed for each of the 96 with-context stimuli. The aim was to produce sentences whose means were less than 2 for the highly constraining long-context condition and between 3 and 4 for the less constraining short-context condition. The sentences were adjusted and retested until these criteria were met.

The final versions of the sentences were recorded by a male speaker of Scots English. These stimuli were then digitized using a PDP 11/40 computer. Visual and auditory information were used to locate four points in each stimulus: the onset of the first clause, the onset of the second clause, and the onset and offset of the test word. From a single recorded sentence, three stimuli were produced. The long-context stimulus was measured from the onset of the first clause to the offset of the test word. (Example: *When I went to see the doctor, he was in the surgery.*) The short-context stimulus was measured from the onset of the second clause to the offset of the test word. (Example: *He was in the surgery.*) The no-context stimulus consisted of the test word alone. (Example: *surgery.*)

The test words were gated automatically by 50-msec increments and were generated afresh at every experimental session by the PDP 11/40 computer. Each subject was presented with one of three experimental sequences. Although every subject heard every test word, no subject heard any test word in more than one context condition.

Procedure

Individual subjects were seated before a computer terminal in a soundproofed room; the stimuli were presented through earphones. The subject was told to respond aloud, after each presentation of a stimulus, with the identity of the gated word, which in the case of the with-context stimuli was the final word in the sentence. The subject was aware that the test word was the last word in a well-formed sentence. These responses were recorded and the resulting tape was used to check the written record of responses made by the experimenter at the time of the experiment. In addition, the subject was to press a key to indicate his/her confidence in his/her responses on a 1–5 scale (1 for *very unsure*, 5 for *absolutely certain*). The subject's keypress triggered the presentation of the next test item. Three practice items preceded the experimental stimuli, which were grouped in blocks according to the context variable; the order of presentation of blocks was counterbalanced. The subject took a break after the first or the second block, according to individual preference. An appropriate visual display appeared on the computer terminal 1 sec before the presentation of a fresh stimulus and .5 sec before the onset of each successive gate.

RESULTS AND DISCUSSION

For each dependent variable, two sets of analyses are presented. In the first, only those responses offered at the first gate (50 msec) are included; the second analysis refers to the responses offered at gates one, two, and three (50, 100, and 150 msec).

Grammatical Category of Responses

It was pointed out above that subjects were informed that the gated word was the last word in the sentences heard in the with-context conditions. Since the gated word was immediately preceded by a determiner (*a* or *the*), its only possible word class was "noun." Therefore, in the first analysis, the dependent variable was the percentage of valid responses that numbered "noun" among their possible word classes. Following Clark (1973), data were submitted to two analyses to yield by-subjects and by-materials values of F. These two values were then used in the calculation of minF'. At the first gate (50 msec), 73% of responses offered in the no-context condition were potential nouns, compared with 83% in the short-context condition and 88% in the long-context condition. This result was significant [$F_1(2,70) = 19.20$, $p < .001$; $F_2(2,94) = 10.18$, $p < .01$; min$F'(2,160) = 6.65$, $p < .01$]. Post hoc analysis of the means by Sheffé test showed that all three means differed at the $p < .01$ level of significance (by subjects). When data collected during the first three gates were examined, a similar pattern emerged: 74% of responses in the no-context condition were nouns, compared with 84% and 93% in the short- and long-context conditions, respectively [$F_1(2,70) = 46.16$, $p < .001$; $F_2(2,94) = 13.43$, $p < .001$; min$F'(2,140) = 10.40$, $p < .001$]. Once again, all three means differed significantly ($p < .01$, by-subjects Scheffé test).

These results fail to support the hypothesis that no context effects will be detected during the processing of the initial portion of a word; on the contrary, when syntactic constraints were present, subjects were actively using them to limit the types of word candidates considered, even after hearing only the initial 50 msec of the acoustic input.

Semantic Appropriateness of Responses

Semantic appropriateness of responses was assessed in the following manner: For each of the two comparisons (no-context vs. short-context and no-context vs. long-context) a separate pair of judges was asked to evaluate the appropriateness of each response made in the early gates of each condition in relation to the sentence introduction heard in the with-context condition. That is, for the first analysis (no-context vs. short-context), no-context responses and short-context responses were evaluated in relation to the short-context sentence, whereas in the second analysis (no-context vs. long-context), no-context and long-context responses were evaluated in relation to the long-context sentences.

Nineteen percent of no-context responses and 65% of short-context responses made at the first gate were judged to be appropriate completions of the short-context sentence fragment. This difference is highly significant [$F_1(1,35) = 338.39$, $p < .001$; $F_2(1,47) = 111.87$, $p < .001$; min$F'(1,73) = 84.08$, $p < .001$]. When data from the first three gates were analyzed, the means rose to 25% for the no-context condition and 74% for the short-context condition [$F_1(1,35) = 562.96$, $p < .001$; $F_2(1,47) = 168.35$, $p < .001$; min$F' = 129.60$, $p < .001$].

The effect was even more marked when the no-context and long-context responses were compared. Only 3% of no-context responses offered at the first gate were appropriate completions of the long-context sentence, compared with 69% of long-context responses [$F_1(1,35) = 710.08$, $p < .001$; $F_2(1,47) = 371.30$, $p < .001$; min$F'(1,80) = 243.81$, $p < .001$]. Once again, these means rose when the data from the first three gates were considered, to 5% and 76% for the no-context and short-context responses, respectively [$F_1(1,35) = 905.42$, $p < .001$; $F_2(1,47) = 669.12$, $p < .001$; min$F'(1,82) = 384.77$, $p < .001$].

In all cases, the improvement in the with-context mean over the no-context mean was significant at a level well in excess of $p = .01$; the subjects were once again clearly making use of contextual cues during their processing of the initial portion of words. This effect is apparent even when the analysis is restricted to those responses produced after hearing only 50 msec of the input.

CONCLUSIONS

In this paper I set out to reexamine the cohort model's prediction that hearers would make no use of context during their processing of the initial portion of a word. For dependent variables relating to both syntactic and semantic constraints, the no-context data were used as a control condition against which subjects' use of contextual information in the with-context conditions was evaluated. In no analysis was the cohort model's prediction borne out: in each case, the number of contextually appropriate responses offered in with-context conditions significantly exceeded the number of appropriate responses offered when no contextual constraints were present. Such results are more consistent with the predictions of a model such as the logogen model (Morton, 1969), which allows top-down and bottom-up information to interact at any stage during the processing of a word.

However, if we accept that hearers *can* make use of top-down information even during the earliest stages of processing, we must ask why they sometimes fail to do so. In Tyler's (1984) experiment, 10% of responses were syntactically inappropriate and 6% were semantically inappropriate. The explanation offered above for this finding was that on at least some occasions subjects were misinterpreting these contextual cues. The data produced in the present experiment indicate that this was indeed the case. For example, most of the subjects who heard the sentence fragment

1(a) When John threw his racquet in the air, it hit a...

followed by the initial 50 msec of the test word *spectator* appeared to have misinterpreted it as

1(b) When John threw his racquet in the air, it hit us...

since they offered responses consistent with such a parse (e.g., *hard*, *on the head*). The same phenomenon was observed in the short-context condition. This erroneous interpretation generally persisted for several gates. Similarly, the sentence introduction in 2(a) (and its corresponding short-context introduction) seem to have been interpreted by most subjects as that in 2(b):

2(a) Although he sometimes watches our team, Bill is not a ... (test word *supporter*)

2(b) Although he sometimes watches our team, Bill is not as ... (example of response: *clever*).

The utilization of left context in word recognition presupposes the accurate recognition of the words of which that context is composed. In the case of the examples just described, the subjects failed to realize that there was a word boundary between the determiner and the test item. Although some writers (e.g., Gimson, 1980; Nakatani & Dukes, 1977) have identified word boundary cues in connected speech, such cues are not consistently present, especially in less formal speech. Similarly, phonological reduction, which might be expected to complicate the task of word recognition, is more prevalent in casual than in formal speech (Brown, 1977). Although hearers make use of contextual cues as they perceive them, their perceptions are not always correct, even when the speech input is very carefully articulated. In casual speech, in which words are often less than 50% intelligible on the basis of the acoustic signal alone (Pollack & Pickett, 1963), hearers are presumably forced to rely increasingly on contextual cues; yet it seems that such cues will not infrequently lead them astray.

REFERENCES

Brown, G. (1977). *Listening to spoken English*. London: Longman.

Clark, H. H. (1973). The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning & Verbal Behavior*, **12**, 335-359.

Forster, K. I. (1981). Priming and the effects of sentence and lexical contexts on naming time: Evidence for autonomous lexical processing. *Quarterly Journal of Experimental Psychology*, **33A**, 465-495.

Francis, W. N. & Kučera, H. (1982). *Frequency analysis of English usage: Lexicon and grammar*. Boston: Houghton Mifflin.

Gimson, A. C. (1980). *An introduction to the pronunciation of English* (2nd ed.). London: Edward Arnold.

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

Marslen-Wilson, W. (1987). Functional parallelism in spoken word recognition. *Cognition*, **25**, 71-102.

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

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

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

Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, **62**, 714-719.

Pollack, I., & Pickett, J. M. (1963). The intelligibility of excerpts from conversation. *Language & Speech*, **6**, 165-171.

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

NOTE

1. There has been considerable debate over the nature of this matching process. Although earlier versions of the cohort model required a one-to-one match between input and lexical entry, in the most recent formulation Marslen-Wilson (1987) appealed to the notion of differential activation of candidates on the basis of some metric of goodness of fit. This aspect of the cohort model will not be discussed in this paper.

(Manuscript received July 21, 1987; accepted for publication November 18, 1987.)