

## Chapter 11

Auditory Lexical Access: Anne Cutler  
Where Do We Start?

The lexicon, considered as a component of the process of recognizing speech, is a device that accepts a sound image as input and outputs meaning. Lexical access is the process of formulating an appropriate input and mapping it onto an entry in the lexicon's store of sound images matched with their meanings.

This chapter addresses the problems of auditory lexical access from continuous speech. The central argument to be proposed is that utterance prosody plays a crucial role in the access process. Continuous listening faces problems that are not present in visual recognition (reading) or in noncontinuous recognition (understanding isolated words). Aspects of utterance prosody offer a solution to these particular problems.

### **1 Lexical Access in Continuous Speech Recognition**

There is, alas, no direct auditory equivalent of the little white spaces that so conveniently segment one word from another in continuous text. Speech is truly continuous. There are, admittedly, certain phonological cues to segmentation; for instance, the phonetic sequence [tb] must span a syllable boundary. But equally there are phonological effects that mask boundaries; the sequence [dj] can be affricated, irrespective of whether it belongs to one word (as in British English *duty*) or two (as in *did you*). There is no reliable cue marking every word boundary in speech, as there is in text.

Thus, in continuous auditory recognition there is a problem of *segmentation* that is absent in visual recognition (where spaces between words provide explicit word-boundary markers) and in the recognition of words in isolation. Segmentation is necessary because lexical access must operate with discrete units, since the lexicon must store discrete units. The number of potential utterances is infinite; no recognizer could possibly store in memory, for eventual match against a possible future input, every complete

utterance that might conceivably be presented. Only the discrete meanings from which utterances are combined can reasonably be stored. When a recognizer is presented with continuous speech, therefore, it cannot begin the process of lexical access until it has taken some decision about how the stream of continuous speech should be segmented into units that one might reasonably expect to be matched in the lexicon.

The segmentation issue is closely connected with the question of the formulation of the lexical access code, i.e., the precise nature of the input to the lexicon. For instance, suppose that input representations in the lexicon were in the form of minimally normalized acoustic templates. Computation of the lexical access code would be trivial; however, attempts at lexical access would have to be initiated at vastly many arbitrarily determined points, leading to a huge imbalance of wasted versus successful access attempts. At the other extreme, suppose that the access code were a string of syllables. Then one might simply assume each syllable boundary to be a word boundary. This strategy would successfully detect each word boundary at relatively little cost in false alarms, but the computational cost of deriving the access code representation would be considerable—especially in English and other stress languages, where syllable boundaries are frequently unclear. Moreover, there is evidence that syllabic segmentation is not used in the recognition of English words (Cutler, Mehler, Norris, and Segui 1986).

The present chapter proposes a strategy for segmentation of continuous speech and for access of lexical forms in languages like English. The strategy is based on prosody. However, as the next section will describe, prosody may not be exploited in the most superficially plausible way.

## 2 Lexical Prosody and Metrical Prosody

Words and sentences both have prosodic structure. Word prosody, in lexical-stress languages such as English, consists of stress pattern: What is the stress level of each syllable of a word? Take, for instance, the words *generate*, *general*, and *generic*. All have different lexical prosody. *Generate* and *general* both have primary stress on the first syllable; *generic* is stressed on the second syllable. However, *generate* has secondary stress on the third syllable and hence is further distinguished from *general*, which has no secondary stress.

The prosodic structure of sentences embraces intonation contour and rhythm. The rhythm of an utterance is the pattern of strong and weak syllables in the words that make up the utterance. Metrical prosody is

another name for this system. Thus, metrical prosody is in one sense a simpler system than lexical prosody, since it has only two levels: strong and weak. A strong syllable is any syllable containing a full vowel; a weak syllable contains a reduced vowel (usually a schwa). The citation form of *generate*, in terms of metrical prosody, has the pattern strong-weak-strong; the citation form of *general* is strong-weak-weak; and the citation form of *generic* is weak-strong-weak. Thus, although in terms of lexical prosody *generate* and *general*—both having initial stress—might seem more like each other than like *generic*, in terms of metrical prosody it is reasonable to consider *general* and *generic*—with only one strong syllable each—as in a sense more like each other than like *generate* (which has two strong syllables).

However, the difference between lexical prosody and metrical prosody is not just that the metrical prosodic system allows only two levels whereas the lexical prosodic system has more. The important difference is in the domain of each system. Lexical prosody, as the name suggests, refers to words. Lexical stress patterns are defined upon canonical pronunciations of words. Importantly, lexical stress patterns of words in citation-form pronunciations may not always be fully realized in actual utterances. (*General*, for example, is often pronounced with two syllables.) The metrical prosodic system, on the other hand, refers to the rhythmic pattern of longer stretches of speech. Again, citation-form rhythms may differ from the rhythms of the same forms produced in conversational context; however, actual rhythms and citation-form rhythms can equally well be described as a sequence of strong and weak syllables.

It might seem most likely that the relevant dimension of prosody for lexical representation, and hence lexical access, should be lexical prosody. As the next sections will demonstrate, such an assumption seems not to be justified.

### 3 Does Lexical Prosody Play a Role in Lexical Access?

In order to know the lexical prosodic structure of a word, the recognizer must know how many syllables the word has; in order to know how many syllables a word has, it is necessary to know where the word begins and ends. This dependence in itself suggests that lexical prosody may not be a crucial component of the lexical access code, at least in continuous speech recognition. If lexical prosodic information is useful for word recognition, then prior awareness of such information should facilitate lexical access; Cutler and Clifton (1984), however, found that lexical decision responses

were not in fact facilitated when subjects were given prior knowledge of stress via a grouping of materials according to stress pattern.

Some pairs of words differ in stress pattern but are otherwise pronounced identically. *FOREarm* and *foreARM*, for instance, each have full vowels in both syllables, despite their stress-pattern opposition. If lexical prosody is entirely irrelevant to the lexical access code, then identical codes should be computed for each member of such pairs. That is, *FOREarm* and *foreARM* should, before the lexical-access stage of recognition, be effectively homophonous. And they are: Cutler (1986) tested such pairs in a cross-modal priming study. In this case the cross-modal priming task, developed by Swinney (1979), served as a diagnostic test for homophony. Swinney showed that homophones prime associates to each of their meanings; when a subject hears the homophone *bug*, for instance, lexical-decision responses to a simultaneously presented visual probe are faster if the probe is related to either the "insect" or the "listening device" meaning of *bug*. That is, both *ant* and *spy* are responded to faster than the matched control word *sew*. Similarly, when a subject hears a sentence containing either *FOREarm* or *foreARM*, in both cases the subject responds to visual probes related to either word (*elbow* versus *prepare*) faster than to matched control words. That is, pairs of words differing only in lexical prosody—such as *FOREarm* and *foreARM*—behave just like other homophones. Presentation of the sound sequence representing either one leads to access of the lexical representation of both. This result is not due simply to partial priming via partial phonetic overlap, since phonetically similar words do not prime one another (Słowiaczek and Pisoni 1986). The homophony of *FOREarm* and *foreARM* shows that differences of lexical prosody alone are not sufficient to produce differences in the lexical access code. Lexical prosody—i.e., the fact that *FOREarm* is stressed on the first syllable and *foreARM* on the second—appears to be irrelevant in lexical access.

This conclusion offers a way out of what would appear to be a dilemma produced by apparently contradictory results from investigations of a related phenomenon, namely the effects of erroneous lexical prosody. Some studies have suggested that misstressed words are harder to recognize. For example, Bansal (1966) found that Indian English, in which stress is signaled in a manner that is unconventional to British English ears, led British English listeners to misperceive stress placement and consequently to misinterpret the speaker's utterance so as to conform with the erroneous stress pattern—even though the chosen interpretation at times conflicted with the phonetic segmental structure of the utterance. Thus, one speaker pronounced *yesterday* with a pitch peak on the second syllable, and sig-

naled the placement of lexical stress on the first syllable by lengthening only. In British English, pitch peaks and lengthening co-occur in the signaling of lexical stress. Listeners reported hearing *or study* rather than *yesterday*; that is, the stress was perceived as occurring where the pitch peak occurred, and the utterance was interpreted as one with lexical stress properly on that syllable even though this required interpretation of the reduced vowel [a] as the full vowel [A]. Similarly, Bond and Small (1983) found that misstressed words were infrequently restored in shadowing; with the number of times a mispronounced word was neither repeated verbatim nor restored to its proper form used as an index of disruptiveness of the mispronunciation, misstressing proved about three times as disruptive as mispronunciation of a single vowel or consonant. Cutler and Clifton (1984), using a semantic-categorization task, also found that misstressed words were disruptive; they were responded to more slowly than their correctly stressed versions.

The latter study, however, also made an explicit comparison between the prosodic and segmental attributes of stress. Cutler and Clifton compared the effects of misstressing words having two full syllables (such as *canteen* or *turbine*) against the effects of misstressing words with one weak syllable (such as *lagoon* or *wallet*). In the latter case, misstressing necessarily resulted in a segmental change as well (i.e., a change from a schwa to a full vowel), whereas in the former case it did not. Only in the latter case did misstressing necessarily inhibit recognition—*LAGoon* and *wALLEt* were responded to significantly more slowly than *laGOON* and *WALLEt*, respectively. But when no segmental change was involved, the effect of the misstressing depended on the direction of the stress shift. Rightward shifts were harmful; *turBINE* was harder to recognize than *TURbine*. Leftward shifts, however, were harmless; recognition of *CANteen* and *canTEEN* did not differ significantly.

Taft (1984) also failed to find effects of misstressing on phoneme-monitoring response time. Detection of phoneme targets was not significantly slowed if the word immediately preceding the target was misstressed. In fact, responses were actually faster if the misstressing involved a leftward shift; *CHAMpagne* produced faster responses than the correct version *chamPAGNE*.

The apparent contradiction between these results disappears, however, with consideration of the distinction between lexical prosody and metrical prosody. All misstressings necessarily alter lexical prosodic structure. As Cutler's study of cross-modal priming showed, however, lexical prosody is irrelevant in lexical access. Only some misstressings appear to inhibit

lexical access. Only some misstressings alter metrical prosody as well as lexical prosody—specifically, when a full vowel is reduced or a reduced vowel becomes full, the metrical structure of the word is altered.

The overall result of the misstressing studies can be summarized as follows: Changes in metrical prosodic structure necessarily inhibit recognition; changes only in lexical prosodic structure do not. Thus, Bansal's Indian speakers were *perceived* as giving full vocalic value to vowels that in British English would normally be reduced; Bond and Small's misstressings all resulted in reduced vowels becoming full; and Cutler and Clifton's *lagoon-wallet* condition similarly produced full vowel quality where reduction was expected. All these misstressings therefore altered the metrical prosody (via alteration of the vocalic segments), and they all produced significant decrements in recognition performance.

Taft, on the other hand, manipulated lexical prosody without altering metrical prosody. The words she used (e.g., *afghan* and *champagne*) had full vowels in both syllables, so shifting stress from one syllable to another left vowel quality intact and the metrical structure unchanged. Cutler and Clifton's *canteen-turbine* condition also left metrical structure intact. Taft found no effect of lexical prosodic shift under these conditions; Cutler and Clifton found no effect for leftward shifts. Rightward shifts, it is true, did inhibit recognition in Cutler and Clifton's study. However, they included several words like *whisky*, with an open final syllable. Bolinger (1981) suggests that the second syllable of such a word is in fact metrically weak. If that is the case, then rightward shifts in the Cutler-Clifton study may actually have altered metrical structure (whereas leftward shifts did not). Thus, the picture may be perfectly consistent: Misstressing only inhibits recognition when metrical prosody is changed, not when only lexical prosody is changed.

This implies that the role of prosody in the lexical access process is by no means a direct one. It is possible to describe the prelexical representations computed for lexical access purely in segmental terms; lexical prosody does not need to be marked in these codes. This is fully consistent with a view of lexical representations such as that proposed by Bolinger (1981), in which lexical entries have no stress patterns but have only segmental representations (in which full vowels are represented as full and reduced vowels as reduced) plus a marker indicating which syllable should receive primary accentuation in citation form. An accurate segmental representation will be all that is needed to access a lexical entry. Reducing a full vowel or giving full value to a reduced vowel (even when, as in the second syllable of *whisky*, the reduced form can be described as a very short version of the

full form) results in an inaccurate segmental representation and hence in poorer recognition performance. The prosodic structure of words is not coded for lexical access; only the segmental structure is relevant.

However, the importance of the distinction between full and reduced vowels in this prelexical segmental representation suggests an indirect role for prosody in the lexical access process. That is, it may be reasonable to claim that there is a sense in which metrical prosody plays a role in lexical access even if lexical prosody does not. Some specific investigations of this issue will be considered in the next section.

#### **4 Does Metrical Prosody Play a Role in Lexical Access?**

Speakers assume that metrical prosody is more important to listeners than lexical prosody. This conclusion can be drawn from further consideration of misstressings. Sometimes speakers misstress a word quite by accident. Here are four examples of such errors in lexical stress (from Cutler, 1980):

- (1) ... from my PROsodic-proSODic colleagues.
- (2) ... each of these acoustic property detectors perhaps being SUBJECT—perhaps being SUBject to ...
- (3) You think it's sarCASm, but its not.
- (4) We're only at the early stages of it, we're still enTHUSiastic.

Sometimes the misstressing produces a change in vowel quality and hence an alteration in metrical prosody—as in (1), where the reduced vowel in the first syllable became a full vowel in the error, or as in (2), where the full vowel in the first syllable was reduced in the error. But sometimes a misstressing produces no metrical change at all—as in (3) and (4), where stress has shifted from one strong syllable, or full vowel, to another.

Speakers do not always correct their slips of the tongue; in fact, the correlation rate for lexical-stress errors (34 percent in the author's corpus) is noticeably lower than the mean correction rates for phonemic (75 percent) or lexical (57 percent) slips cited by Nootboom (1980). When speakers do correct lexical-stress errors, however, they correct them significantly more often if the metrical prosody has been altered (61 percent) than if only the lexical prosody has changed (21 percent) (Cutler 1983). That is, errors like (1) and (2) are far more likely to be corrected than errors like (3) and (4). Speakers appear *to* assume that changes in metrical prosody will threaten listeners' reception of the message more than changes in lexical prosody.

Table 1

Slips of the ear: juncture misperception.

Spoken	Perceived
it was illegal	it was an eagle
assistant	his sister
a Coke and a Danish	a coconut Danish
my gorge is...	my gorgeous
she's a must to avoid	she's a muscular boy
fornication	for an occasion
paint your ruler	paint remover

Suppose, however, that a slip is made by a listener. Studies of slips of the ear show that prosodic misperceptions are very rare indeed. Metrical prosody, particularly, is resistant to distortion—the parts of the speech signal that are least susceptible to distortion are the vowels in stressed syllables (Bond and Garnes 1980; Browman 1978). The only way in which metrical prosody is distorted in slips of the ear is that weak syllables may be lost or duplicated.

Table 1 presents some slips of the ear. They are all errors in which juncture has been misperceived—that is, word boundaries have been added, lost, or shifted. It can easily be seen that, as with slips of the ear in general, metrical structure is preserved—strong syllables are perceived as strong, weak ones as weak. But what is interesting about these slips is the direction of the boundary mislocations: Boundaries tend to be perceived at the onset of strong syllables rather than weak. When "she's a must to avoid" is perceived as "she's a muscular boy", a boundary has been added prior to the final strong syllable, while boundaries before the two weak syllables preceding it have both been deleted. Similarly, when "a Coke and a Danish" is perceived as "a coconut Danish", boundaries before two weak syllables have been deleted, and when "it was illegal" is perceived as "it was an eagle", a boundary has been moved from the onset of a weak syllable to the onset of a strong syllable.

Sometimes the reverse is true, as in the final two examples in the table. Perception of "paint your ruler" as "paint remover", for instance, deletes a boundary before a strong syllable. However, such examples are in the minority. Sally Butterfield and I examined all the juncture misperceptions we could find in my own collection of slips of the ear and in published collections such as Bond and Games 1980 and Browman 1978. More than two-thirds of them (roughly the proportions in the table) conformed to



the generalization that boundaries were perceived at the onset of strong syllables rather than weak ones. The effect was statistically significant (Butterfield and Cutler 1988).

This suggests a specific way in which metrical prosody may be exploited in lexical access: Strong syllables may be taken to be word onsets. Thus, metrical prosody may provide a way of dealing with the crucial segmentation problem in continuous speech recognition. Some further recent work from our laboratory offers strong support for this hypothesis.

Cutler and Norris (1988) investigated the detection of words embedded in nonsense matrices with differing metrical structures. For example, the word *mint* was embedded either in *mintayf* [mintef], in which the second syllable was strong, or *mintef* [mintef], in which the second syllable contained a schwa (i.e., was weak). Cutler and Norris reasoned that, if segmentation were guided by metrical prosody such that boundaries were hypothesized prior to each strong syllable, then *mintayf* would be segmented (min-tayf), whereas *mintef*, with a weak second syllable, would not. A word spread across two segmentation units should prove more difficult to detect than the same word in an unsegmented string. Indeed, detection of *mint* in *mintayf* was significantly slower than detection of *mint* in *mintef*. Further experiments ruled out several confounds. For instance, it was not the case that *mint* in *mintayf* was spoken in such a way that it was less like some canonical lexical template for *mint* than *mint* in *mintef*. Cutler and Norris demonstrated this by presenting subjects with the same strings from which the final vowel-consonant sequence had been removed. If the actual articulation of the *mint* token were responsible for the difficulty of detecting *mint* in *mintayf*, then *mint* from *mintayf* should still be harder to recognize than *mint* from *mintef*. However, both tokens were detected equally rapidly.

Again, it was not the case that the greater difficulty of detecting *mint* in *mintayf* than in *mintef* was simply due to the nature of the second syllables. It could be that subjects waited till the end of the item to initiate a response, and they had to wait longer in *mintayf* than in *mintef* because the second syllable is longer. Alternatively, it could be that the louder a second syllable is, the more it interferes with processing of the first syllable, so that the second syllable of *mintayf*, being louder than the second syllable of *mintef*, interfered more with the processing of *mint*. But Cutler and Norris disposed of these possibilities by comparing detection of *mint* in *mintayf* versus *mintef* with detection of *thin* in *thintayf* versus *thintef*. The second syllable of *thintayf* is just as much longer and louder than the second syllable of *thintef* as the second syllable of *mintayf* versus *mintef*. If simple loudness or duration of the second syllable were responsible for the difficulty of *mint*

in *mintayf* versus *mintef*, then detection of *thin* in *thintayf* should similarly be harder than detection of *thin* in *thintef*. On the other hand, if Cutler and Norris' interpretation of the difficulty of detecting *mint* in *mintayf* were correct, namely that it was due solely to the difficulty of detecting a word spread over two segmentation units, then detection of *thin* should be equally easy in both *thintayf* and *thintef*; although *thintayf*, with a strong second syllable, would be segmented (*thin-tayf*), the word *thin* belongs only to the first segmentation unit, so that segmentation should not hamper its detection in any way. Cutler and Norris' prediction was supported: Although *mint* was again harder to detect in *mintayf* than in *mintef*, *thin* was detected equally rapidly in both *thintayf* and *thintef*.

The conclusion drawn by Cutler and Norris was that metrical prosody is indeed exploited in word recognition. Specifically, it forms the basis of a strategy of segmentation, whereby boundaries are postulated at the onset of strong syllables.

## 5 Metrical Prosody of the English Vocabulary

The rationale for positing boundaries prior to strong syllables, it may be assumed, is that such boundaries are likely to be lexical unit (i.e., word) boundaries. Taft (1984) has direct evidence in support of this suggestion. She presented listeners with phonetically ambiguous strings, such as *lettuce/let us* (which is metrically strong-weak) and *invests/in vests* (which is metrically weak-strong). For the strong-weak strings, listeners greatly preferred the one-word interpretation; two-word interpretations were chosen more often for the weak-strong strings. That is, boundaries were inserted prior to strong rather than weak syllables. Thus, English listeners appear to segment speech on the working hypothesis that words will begin with strong syllables. However, this strategy will obviously not succeed with all words. In many English words the first syllable's vowel is weak—*appear*, *begin*, and *succeed* are three examples from the present paragraph. Why should listeners adopt a strategy that may often fail?

Closer consideration of the characteristics of the English vocabulary, however, suggests that a working hypothesis that words begin with strong syllables will fail surprisingly seldom in the recognition of everyday spoken English. First, there are in fact many more words beginning with strong than with weak syllables; second, words beginning with strong syllables have a higher frequency of occurrence than words beginning with weak syllables.

All the most common lexical prosodic patterns in English have a full vowel in the first syllable (Carlson, Elenius, Granstrom, and Hunnicutt 1985). However, more detailed information is available in the metrical statistics which David Carter and I compiled from a 30,000-word dictionary of British English. Seventy-three percent of all words (and 70 percent of polysyllabic words) were listed with a phonetic transcription in which the first vowel was full. (A 20,000-word corpus of American English shows an almost identical distribution: 78 percent of all words, and 73 percent of polysyllabic words, begin with a strong syllable.<sup>1</sup>) However, in a subset of the larger British English corpus, consisting of the 13,000 most common words, it was possible to examine the metrical structure of the vocabulary as a function of word class and frequency of occurrence. We assumed that all monosyllabic closed-class (grammatical) words, irrespective of their phonetic transcription in the dictionary, would be metrically weak in continuous speech. Using this assumption, we found that 72.32 percent of the whole of this 13,000-word subset, and 73.46 percent of the open-class (lexical) words, consisted of or began with strong syllables. But when the mean frequency with which each type of word occurs is taken into account, about 85 percent of open-class words in average speech contexts have full vowels in their (only or) initial syllables.

Of course, many of the words in an average utterance will be grammatical words, such as determiners, conjunctions, and pronouns, and nearly all of these will be monosyllabic and metrically weak. The mean frequency of grammatical words is very high indeed. The proportions of strong and weak onsets are exactly reversed in comparison with the open-class case: Only about 25 percent of grammatical words are polysyllables with strong onsets. (Another 25 percent are polysyllables with weak onsets; the remaining 50 percent are monosyllabic.) Thus, of all words in the average utterance, both grammatical and lexical, it may be that only a minority will have strong initial syllables. But it is highly debatable whether the process of *lexical access* as it was outlined at the beginning of this chapter applies in the same sense to grammatical words—i.e., whether grammatical words have lexical representations of the same kind as lexical words, and whether the process of converting sound to meaning in speech recognition is of the same nature and complexity for grammatical words (especially those that are monosyllabic and metrically weak) as for lexical words. Meaning itself is, after all, not of the same nature for grammatical as for lexical words; the meaning of grammatical words is context-dependent to a far greater degree (consider, for example, *to* in "to swim", "to Cambridge", "to John", "to arms", and "to a far greater degree").

Whatever the lexical model, however, it is clear that the metrical distribution of the English vocabulary allows listeners to extract considerable information from metrical structure. Wherever there is an open-class word there will be at least one strong syllable, and the likelihood is that open-class words will be or begin with strong syllables.

## 6 The Beginnings of Words and the Beginnings of Lexical Access

The thesis proposed in this chapter has been that the pattern of occurrence of strong and weak syllables offers the basis of a strategy for initiating the lexical access process in the recognition of continuous speech: Start a potential lexical access procedure whenever a strong syllable occurs. If word boundaries were reliably marked in continuous speech, there would be no need to invoke such a strategy. This strategy is tailored to the specific problems of recognition in the auditory modality, of continuous speech rather than isolated words.

As the previous section described, the strategy will successfully locate the onsets of the majority of lexical words in the average communication. The strategy will not, as was also pointed out, locate the onsets of words beginning with weak syllables—*appear*, *begin*, *succeed*, and the like. However, it is at least conceivable that the strategy of treating strong syllables as if they were onsets is supplemented by an ancillary strategy whereby lexical words beginning with weak syllables can, under appropriate circumstances, be successfully accessed via their strong syllables. That is, may not the phonetic strings [pie], [gIn], and [sid] serve as one potential access code for, respectively, the words *appear*, *begin*, and *succeed*? In fact, precisely such a model of lexical access for this type of word has been postulated independently by several authors (Cutler 1976; Bradley 1980; Grosjean and Gee 1987). In such a model, the lexical access process for words like *appear* might be somewhat more complicated than the lexical access process for lexical words that actually do begin with strong syllables. There is evidence that this may indeed be so. This evidence comes from recent work with the gating task (Grosjean 1980), in which successively larger fragments of a word are presented to listeners, who are asked to attempt to identify the word. Studies using this task have investigated how much of a word must actually be heard before the listener can be reasonably sure of what the word is. When words are presented in isolation, recognition of words with strong versus weak initial syllables is not significantly different (William Marslen-Wilson, personal communication). This is also true of the recognition of words in continuous speech, but only if the speech

has been carefully read; if natural spontaneous speech is recorded and used in a gating task, then recognition of words with strong first syllables is significantly facilitated in comparison with recognition of words with weak first syllables (MacAllister, in preparation).

The one problem with postulating lexical access via strong syllables for words that begin with weak syllables is that there exists considerable evidence that auditory lexical access proceeds "left to right"—i.e., that the beginnings of words are always accessed before later portions. This evidence underlies the cohort model of auditory word recognition (see, e.g., Marslen-Wilson 1980 and this volume), in which, for instance, a word-initial phonetic string [pie] will activate the words *pierce*, *peerage*, *pianist*, etc. (but not *spear*, not *impious*, and not *appear*).

However, there is a conflict only if one assumes that lexical access must always be based on one and the same access code. This is of course not true of lexical access in language production (i.e., of the access of sound via meaning); any of a number of semantic specifications (e.g., "the originator of the cohort model", "Lolly Tyler's husband", "the editor of this book", "that tall chap in Room 155") will suffice to call up a particular sound pattern (e.g., [wɪljəmzɛnzwɪlsən]). In lexical access in perception it is certainly possible that a given meaning could be accessed via any one of several alternative access codes (e.g., the phonetic strings [pie] and [epie]), and if the starting point of a lexical string is specified it may well be optimally efficient to process it strictly left to right irrespective of how it begins. However, as was stressed at the beginning of this chapter, the major problem for lexical access in natural speech situations is that word starting points are *not* specified. The evidence presented here has shown how prosodic structure, in particular metrical prosodic structure, can offer a way out of this dilemma. Where do we start lexical access? In the absence of any better information, we can start with any strong syllable.

### Acknowledgments

Some of the work described in this chapter was supported by grants from the Alvey Directorate and British Telecom. Thanks to the colleagues with whom collaborative work described here was performed: Sally Butterfield, Chuck Clifton, and Dennis Norris. Thanks to Carol Fowler, William Marslen-Wilson, and Dennis Norris for helpful comments on the manuscript.

### Note

1. I am grateful to Uli Frauenfelder for compiling the American statistics.

## References

- Bansal, R. K. 1966. The Intelligibility of Indian English. Ph.D. thesis, London University.
- Bolinger, D. L. 1981. Two Kinds of Vowels, Two Kinds of Rhythm. Indiana University Linguistics Club, Bloomington.
- Bond, Z. S., and Games, S. 1980. Misperceptions of fluent speech. In *Perception and Production of Fluent Speech*, ed. R. Cole. Hillsdale, N.J.: Erlbaum.
- Bond, Z. S., and Small, L. H. 1983. Voicing, vowel and stress mispronunciations in continuous speech. *Perception and Psychophysics* 34: 470-474.
- Bradley, D. 1980. Lexical representation of derivational relation. In *Juncture*, ed. M. Aronoff and M.-L. Kean. Saratoga, Calif.: Anma Libri.
- Browman, C. P. 1978. Tip of the Tongue and Slip of the Ear: Implications for Language Processing. UCLA Working Paper in Phonetics 42.
- Butterfield, S., and A. Cutler. 1988. Segmentation errors by human listeners: Evidence for a prosodic segmentation strategy. In *Proceedings of SPEECH '88: Seventh Symposium of the Federation of Acoustic Societies of Europe*. Edinburgh.
- Carlson, R., Elenius, K., Granstrom, B., and Hunnicutt, S. 1985. Phonetic and orthographic properties of the basic vocabulary of five European languages. *Speech Transmission Laboratory (Stockholm): Quarterly Progress and Status Report*, issue 1: 63-94.
- Cutler, A. 1976. Phoneme monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics* 20: 55-60.
- Cutler, A. 1980. Errors of stress and intonation. In *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*, ed. V. A. Fromkin. New York: Academic.
- Cutler, A. 1983. Speakers' conceptions of the functions of prosody. In *Prosody: Models and Measurements*, ed. A. Cutler and D. R. Ladd. Heidelberg: Springer.
- Cutler, A. 1986. *Forbear* is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech* 29: 201-220.
- Cutler, A., and Clifton, C. E. 1984. The use of prosodic information in word recognition. In *Attention and Performance X*, ed. H. Bouma and D. G. Bouwhuis. Hillsdale, N.J.: Erlbaum.
- Cutler, A., J. Mehler, D. Norris, and J. Segui. 1986. The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language* 25: 385-400.
- Cutler, A., and Norris, D. G. 1988. The role of strong syllables in segmentation for lexical access. *Journal of Experimental psychology: Human Perception and Performance* 14: 113-121.
- Grosjean, F. 1980. Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics* 28: 267-283.
- Grosjean, F., and J. Gee. 1987. Prosodic structure and spoken word recognition. *Cognition* 25: 135-155.

- MacAllister, J. Lexical stress and lexical access. Ph.D. thesis, University of Edinburgh (in preparation).
- Marslen-Wilson, W. 1980. Speech understanding as a psychological process. In *Spoken Language Generation and Understanding*, ed. J. C. Simon. Dordrecht: Reidel.
- Nooteboom, S. G. 1980. Speaking and unspeaking: Detection and correction of phonological and lexical errors in spontaneous speech. In *Errors in Linguistic Performance: Slips of the Tongue, Ear, Pen and Hand*, ed. V. A. Fromkin. New York: Academic.
- Slowiaczek, L. M., and Pisoni, D. B. 1986. Effects of phonological similarity on priming in auditory lexical decision. *Memory and Cognition* 14: 230-237.
- Swinney, D. A. 1979. Lexical access during sentence comprehension: (Re)considerations of context effects. *Journal of Verbal Learning and Verbal Behavior* 18:645-659.
- Taft, L. 1984. Prosodic Constraints and Lexical Parsing Strategies. Ph.D. thesis, University of Massachusetts.