



Article scientifique

Article

1985

Published version

Open Access

This is the published version of the publication, made available in accordance with the publisher's policy.

Cross-linguistic approaches to lexical segmentation

Frauenfelder, Ulrich Hans

How to cite

FRAUENFELDER, Ulrich Hans. Cross-linguistic approaches to lexical segmentation. In: Linguistics, 1985, vol. 23, n° 5, p. 669–688. doi: 10.1515/ling.1985.23.5.669

This publication URL: <https://archive-ouverte.unige.ch/unige:84487>

Publication DOI: [10.1515/ling.1985.23.5.669](https://doi.org/10.1515/ling.1985.23.5.669)

Cross-linguistic approaches to lexical segmentation¹

ULI H. FRAUENFELDER

Abstract

This paper examines the problem of lexical segmentation in order to identify language-specific and language-independent properties of lexical processing systems. A potential universal of lexical segmentation is first evaluated for its efficacy across languages. This postlexical segmentation strategy assumes that listeners recognize each word rapidly enough to be able to predict both its offset and the onset of the following word. It is argued that this strategy cannot account for every lexical segmentation decision in utterances in any language. Listeners must also use other types of segmentation information. Two types of segmentation information, distributional and relational, which appear to vary cross-linguistically, are presented with examples. To exploit this information, listeners must infer word boundaries or boundaries of other domains from events or sequences of events in the signal. The problem posed for current models of spoken word recognition by the integration of such segmentation information with that from the putative postlexical strategy is discussed.

Introduction

Our understanding of the mental processes and structures that enable listeners to transform the continuous speech stream into a semantic interpretation has progressed considerably in the last decade. This progress is evident in the study of *lexical processing systems* which listeners use to access and exploit lexical knowledge stored in the mental lexicon. There are now several quite detailed models of spoken word recognition² constructed on the basis of a large number of empirical findings. Unfortunately, these findings have been limited to a handful of relatively similar languages.³ As a consequence, it is still not known how appropriate these models are for structurally different languages. It

remains an open but fundamental question how lexical processing systems vary as a function of the structural properties of their respective languages.

The aim of this paper is to illustrate how cross-linguistic approaches to lexical processing can answer such questions. Two different but complementary approaches can be distinguished. The first approach involves finding shared characteristics of language processing systems. It is reasonable to assume that much of the basic underlying structure of the human language processor does not vary dramatically across languages. General cognitive constraints (for example, memorial and attentional) along with properties of speech (for example, its sequential distribution in time) should ensure that processing systems share some global characteristics.

Some suggestions for language-independent or *core* properties of processing systems have been made by Marslen-Wilson and Tyler (1981). They argue that processing systems are optimal and interactive and have obligatory processes that give priority to bottom-up analyses. Thus, for example, an essential property of the cohort model developed by these authors is *EARLY* word recognition, that is, recognition after an efficient sequential analysis of only the initial part of a word. Properties like these must be tested with cross-linguistic experimentation to determine whether they are truly universal.

The second complementary approach involves looking for cross-linguistic processing differences rather than similarities. A contrastive analysis of the linguistic features (phonetic, phonological, morphological, or prosodic) of languages can translate into predictions of processing differences. These predictions can be tested with experimental cross-linguistic comparisons and eventually generalized to patterns of differences for more languages distinguished by the same feature. Ultimately, underlying higher-order principles that explain these patterns in processing variation must be determined. This approach involves identifying *DIAGNOSTIC* linguistic and phonetic features that correlate reliably with characteristics of the processing system.

This paper aims to show that segmentation processes offer a good starting point for investigations of cross-linguistic similarities and differences in lexical processing systems. First, the assumptions of leading word-recognition models about lexical access and lexical segmentation will be examined. In particular, a strategy of lexical segmentation that is assumed to account for the listener's ability to segment the continuous speech stream will be evaluated. The results of this evaluation show that this strategy alone cannot account for all segmentation decisions; listeners must use other types of segmentation information as well. Two further

types of segmentation information, distributional and relational, will be presented along with examples from different languages. Finally, some suggestions will be made as to how these different sources of segmentation information can be combined in processing models and how these models might differ across languages.

Lexical segmentation in psycholinguistic models

A widely accepted class of spoken word recognition models (Marslen-Wilson and Welsh 1978; Cole and Jakimik 1980) has emerged from research in English and a few other European languages. Such models make several basic assumptions about lexical processing that can be summarized as follows:

1. Listeners recognize words sequentially one at a time.
2. Listeners access words by their beginnings.
3. Listeners recognize words before hearing them completely.
4. Listeners locate the beginning of a word by early recognition of the preceding one.

From a cross-linguistic perspective, one can ask whether a model making these assumptions is equally appropriate for all languages. Proponents of such a model would presumably claim that it does provide a universal description of lexical processing for any language. This claim will be evaluated by focusing on lexical segmentation. More specifically, the adequacy of assumption 4 (above) for lexical segmentation in diverse languages will be examined, and alternative language dependent and/or independent segmentation strategies will be sought.

This general approach to lexical segmentation will be called the *postlexical strategy* of lexical segmentation (Taft 1984). It presupposes that listeners recognize words before completely hearing them (assumption 3). As a result of this early recognition, the 'phonological code' (stored in the lexicon) of the word being processed becomes available soon enough for it to be used to anticipate the end of this word and to predict the beginning of the following one. The earliness of word recognition constitutes an important property of the cohort model, as was mentioned above. In the following section, the factors that determine the time-course of word recognition and the efficacy of the postlexical strategy within this model will be examined and evaluated from a cross-linguistic perspective.

The cohort model proposes that when a listener hears the beginning of a word, a large subset of his/her lexicon containing all lexical entries or candidates that are compatible with this beginning is selected. This subset

or 'cohort' is successively reduced in size as word candidates that mismatch subsequent sensory input drop out. This reduction process continues until only one candidate word remains, at which point this word is recognized, and its phonological code becomes available for lexical segmentation. According to the cohort model, early recognition is possible for a number of reasons. These include the structure and distribution of items in the lexicon, the efficiency of the sequential analysis of the signal, and the effect of higher-order semantic and syntactic context. These factors will each be examined in turn.

The structure of the lexicon imposes strong constraints on the word recognition process. Indeed, the time-course of word recognition, or the speed with which any cohort is reduced to a single member, depends upon the structure and distribution of entries in the lexicon. It is possible to determine, using a representative sample of the language (i.e. a phonetic dictionary), that point in each word at which there are no other words sharing its initial string of phonemes, that is, its *uniqueness point*. For the postlexical segmentation strategy to be effective, a word's uniqueness point must be located before or at its last phoneme; the word must be *segmentally redundant*.

Statistical analyses of computerized phonetic dictionaries provide one way to evaluate the phonological redundancy of a lexicon. Several such studies have been conducted on common European languages. One study (Marslen-Wilson 1984) determined the average number of words in American English that are compatible with initial sequences of different lengths. It was found that the mean number of words for word-initial sequences of three and four phonemes was 1.9 and 1.4 words, respectively. Thus, the uniqueness point for this sample occurred after approximately 3-5 phonemes. Another study of corpora of 10,000 words in five languages (Swedish, English, German, Italian, and French) by Carlson et al. (1985) produced a similar pattern. The median uniqueness point in all five languages was two phonemes less than the median number of phonemes in the words (including a marker for the end of the word). These two studies suggest that a significant proportion of words are segmentally redundant and can be recognized before their ends. However, another analysis (Luce 1984) of an American phonetic dictionary showed that many shorter words are part of morphologically unrelated longer words (i.e. 'car' is part of 'cart' and 'cartoon', etc.) and consequently are not unique before their ends. This analysis revealed that when word frequency was taken into account, the probability that a word diverged from all other words before its end was only 0.39. This finding suggests that up to 60% of all words found in continuous speech are not segmentally redundant in isolation. As a consequence, the postlexical

strategy of segmentation is frequently not of use, at least in the absence of other information.

The second important factor conditioning the time-course of word recognition is the nature of the acoustic-phonetic analysis. This analysis has important consequences for the definition of the uniqueness point and segmental redundancy. A simplifying assumption for defining uniqueness points, made both in the computer analyses presented above and implicitly in the cohort model, is that the acoustic signal is analyzed *sequentially, categorically, and correctly* as a sequence of phonemes.⁴ However, the uniqueness point can be shifted to earlier or later in a word (and the segmental redundancy increased or decreased, respectively), depending upon the way listeners are assumed to process the sensory input.

If listeners can analyze the sensory input more effectively by using additional information, they will recognize words earlier by reducing the cohort more rapidly. For instance, it is well known that a syllable is longer when it is itself a word than when it is only part of a longer word. If rhythmic information such as syllable or phoneme duration can be used by the listener to infer the length of the word being heard, the problem of words within words is partially resolved. The acoustic-phonetic processor with its greater discriminatory power would eliminate longer words from the cohort of a short one and vice-versa. However, the cohort model's estimate of word recognition may not be conservative enough if the listener makes *PROBABILISTIC* rather than categorical decisions on the input. While listeners may be able to make categorical decisions with high quality citation forms, it seems unlikely that they can do so with fluent speech, given its noisy and ambiguous nature. If the processing of the speech input is probabilistic, as some experimental results (Pisoni and Tash 1974; Massaro and Cohen 1983) suggest, then the uniqueness point does not provide an accurate measure of word recognition. In that case, word recognition could only take place *AFTER* the 'uniqueness point', once additional phonetic information allows the intended word to deviate sufficiently from all other words in the language (Marcus and Frauenfelder i.p.). Probabilistic decisions imply later recognition, reduced segmental redundancy, and hence less effective postlexical segmentation.

The cohort model also assumes that successive decisions on the sensory input are always made correctly. Indeed, the misperception of a single segment would have disastrous consequences on lexical processing. In particular, the incorrectly perceived segment would first eliminate the intended word from the cohort, thereby prohibiting its recognition. The failure to recognize this word or the recognition of another incorrect word would lead to a breakdown in lexical segmentation from which recovery could be difficult. The postlexical strategy requires an initial correct

segmentation to start up the assumed cycle of lexical segmentation and lexical access. For cases of breakdowns (or even for the first segmentation decision on an utterance picked up midstream), some other type of segmentation information, for which no mechanism is currently postulated, is required.

Neither the segmental redundancy of words in the lexicon nor the efficiency of the acoustic-phonetic analysis appears sufficient to guarantee the success of the postlexical strategy. This fact places a tremendous burden on the other important source of information that could affect the time-course of word recognition, the preceding sentential context. To make word recognition earlier and postlexical segmentation possible, context must presumably have the power to eliminate contextually inappropriate candidates from the cohort. Despite the many studies of context effects, no generally accepted account of how and when context affects lexical access has been found. It is not possible to do justice here to the controversy on the locus of context effects (but see Tanenhaus and Lucas 1985 and Marslen-Wilson *i.p.*). Nevertheless, there is no conclusive evidence available to demonstrate that contextual constraints are strong enough to always permit the consistently early word recognition required for successful postlexical segmentation.

This theoretical discussion suggests that while the postlexical segmentation strategy can contribute universally to lexical segmentation decisions, it is not sufficient to make all such decisions. Before looking at other possible sources of information that contribute to lexical segmentation, some experimental results from several different languages bearing on this strategy will be presented. It is rather surprising, given the central importance of the lexical segmentation process in these lexical access models, that so little experimental attention has been paid to it. There are, in fact, few if any direct psycholinguistic tests of the postlexical strategy or of any other strategy of lexical segmentation.

The earliness of word recognition constitutes the fundamental requirement for efficient postlexical segmentation. Marslen-Wilson (1984) summarizes results from a variety of experimental tasks in English which converge on extremely short recognition times (of approximately 200 msec) for words in normal contexts. Other results show that the phonological code of a word stored in the lexicon becomes available early. In a phoneme monitoring study with isolated Dutch words (Marslen-Wilson 1984), subjects were asked to detect phoneme targets in different positions within words. The results showed that phoneme detection times decreased steadily as the target appeared later in the word. Further, a strong correlation was found between these RTs and the distance between the target phoneme and the recognition point of the word determined

independently in another study. These results suggest that the stored phonological code becomes available upon word recognition before the end of the word is reached.

Some experimental findings in French (Mehler et al. 1978) indicate that the syllabic length of a word affects the segmentation of the following word. Subjects were instructed to detect word-initial phonemes presented in sentences. Faster phoneme detection times were obtained when the immediately preceding word was long. Since the length of a word is generally correlated with its segmental redundancy, these results provide some indirect evidence for the postlexical segmentation strategy. When subjects process a longer segmentally redundant word, they can use the postlexical strategy more effectively to locate the word boundary and identify the target phoneme more quickly. Finally, there are some studies dealing with the role of sentential context in lexical segmentation. For example, a phoneme-monitoring study (Foss et al. 1980) showed that the transitional probability of the word which precedes the target-bearing word affected the phoneme detection times: the more predictable this word was, the faster phoneme detection times were on the following word. These results suggest that the sentential context assisted lexical access and lexical segmentation, thereby making detection of the word-initial target easier.

These results taken together provide some indirect evidence for the postlexical strategy of lexical segmentation. Other results, however, suggest that the postlexical strategy is not always effective. Grosjean and Gee (1984) have questioned the assumption of efficient sequential access to words as they are heard. In particular, they argued that some words are not recognized until well after their ends. To support this claim they presented experimental evidence obtained using the gating procedure. Subjects were presented successively longer acoustic fragments and were asked to say what word (or words) they had heard a part of. The results revealed that subjects are unable to recognize low-frequency monosyllabic words with little sentential context until they had heard at least the following one or two words. These subjects were clearly not able to use the postlexical segmentation strategy.

Listeners can make consistent segmentation decisions even in instances in which the postlexical segmentation strategy cannot provide segmentation information. Nakatani and Schaffer (1977) found that listeners could correctly segment an utterance into words even if it contained no segmental information. Subjects received 'reiterant speech', that is, a sequence of repetitions of the syllable *ma* that retained stress, rhythm, and pitch. The subjects could use this prosodic information for making lexical segmentation decisions.

Furthermore, Taft (1984) found evidence that listeners segment speech using heuristic strategies. According to one such strategy, strong syllables are taken to be putative word onsets. Her subjects showed strong preferences for segmenting ambiguous words and nonwords according to this strategy. Norris and Cutler (this issue) also suggest that lexical segmentation is guided by a juncture detector that appeals to heuristics similar to those described by Taft. These results taken together suggest that listeners can and do rely on segmentation information other than that coming from the postlexical segmentation strategy.

The present examination of the theoretical assumptions underlying the postlexical strategy and of the experimental results testing these assumptions is not conclusive. While this segmentation strategy is of value in the segmentation of any language, it is apparent that it cannot account for all segmentation decisions in any language. Indeed, the postlexical strategy does not always provide sufficient boundary information and may even given incorrect information. Apparently, listeners must often make lexical segmentation decisions using other segmentation information.

Given the nature of the speech wave, it is perhaps not surprising that psycholinguists have put so much emphasis on the postlexical segmentation approach. It provides an easy solution to a difficult problem. Indeed, it is not obvious how listeners manage to segment fluent speech otherwise since on the one hand, the information corresponding to successive segmentation units like phonemes, syllables, and words often overlaps (coarticulation), or on the other, boundaries between these units are not systematically marked in the signal (Cole and Jakimik 1980; Zue and Schwartz 1980). Cole and Jakimik have even claimed that less than 40% of all word boundaries are marked by physical events (essentially pauses).

This estimate probably falls below the actual number of marked boundaries. While word boundaries are rarely marked by highly salient events like silence, there is other (more subtle?) information that the listener can exploit to detect word boundaries. Phonetic research (cf. Nakatani and Dukes 1977; Lamel 1984) has begun to uncover robust acoustic cues that signal word boundaries. Lamel (1984) found a number of phonetic cues that differentiated the same sequence of phonemes with different word boundaries (e.g. $p\#r$, $\#pr$, and $p\#pr$, where $\#$ represents the word boundary). More such research should lead to the discovery of further acoustic correlates of boundaries. However, such segmentation information probably varies both within and across languages. There are few segmentation cues systematically present in all utterances within the same language or in all languages. This heterogeneity of segmentation information makes it much more difficult to find alternative segmentation strategies as simple as the postlexical segmentation strategy. Nonetheless,

since the postlexical strategy is inadequate by itself, other segmentation information must be used by the listener. In the following section a taxonomy of such information will be presented as a first step toward identifying other segmentation processes.

Distributional and relational segmentation information

It is obvious that any segmentation decision hinges upon the classification of some event or upon the analysis of a sequence of events. It is determining the nature of these events that is critical for understanding segmentation. For the purposes of the present discussion, two main types of segmentation information, *distributional* and *relational*, will be differentiated.

Distributional information⁵ refers to information or events that occur in a fixed position within a higher domain. Thus, a higher or larger domain (like the syllable or the word) constrains the properties of the lower-level units (phones or phonemes) occupying a specific position within this domain. Thus, if listeners identify the relevant property of the lower-level unit, they can then infer the position of that unit within a higher domain, and thereby detect the boundaries of this domain. For example, allophones (or phonemes) that are restricted to word-initial position provide a good example of distributional information for word boundaries; these allophones can signal the location of word onsets to the listener.

Relational segmentation information (similar to the notion of *Gruppen-signale*; Trubetzkoy 1977) refers to a sequence of information or events that can occur either within or across domains. The higher-level domain constrains the sequence of events at a lower level. For example, phonotactic constraints determine where a sequence of phonemes belongs within the higher-order syllabic domain (e.g. two successive fricatives generally belong to different syllables, whereas a vowel usually belongs to the same syllable as the preceding consonant). For lexical segmentation, the presence of two successive stressed syllables might inform the listener that these syllables belong to different words.

Distributional and relational segmentation information can signal not only the PRESENCE but also the ABSENCE of a boundary. For instance, the presence in a language of allophones never found at word boundaries indicates the absence of a word boundary. This type of segmentation information has been referred to as a 'fusion' cue by Nakatani (1980) and 'negative' information by Trubetzkoy (1977) in contrast to the 'fission' cues and 'positive' information described above. Furthermore, it should

be noted that both distributional and relational constraints can be defined in terms of domains (e.g. syllabic and metrical foot domains) other than the lexical domain. The examples of distributional and relational information given in the following section will be restricted to the lexical domain. These examples will be drawn from several different languages and will be organized according to the level that is constrained by the lexical domain.

Distributional segmentation information

i. Acoustic-phonetic/lexical domain

Certain phonetic features like aspiration of consonants and laryngealization of vowels may be found only in word-initial allophones. Trubetzkoy (1977: 244) gives the example of Tamil in which voiceless stops are always aspirated when they occur in word-initial position. Quantitative distributional cues such as the systematic lengthening of word-initial stop consonants also deserve mention.

ii. Phoneme/lexical domain

There are no attested absolute distributional constraints for phonemes such that all words begin (or end) with one specific phoneme that cannot occur elsewhere in the word. There are, however, some weaker distributional constraints. In Bamileke (Hyman 1978), fricative phonemes can occur only at the beginning of words. In English, the absence of words with initial /ʒ/ or /ŋ/ constitutes negative segmentation information.

iii. Syllable/lexical domain

It appears that in both stress-accent and pitch-accent languages, there is at least one primary stressed syllable per word. There are also further constraints on the syllable that affect its internal structure (heavy versus light syllable) or its relative stress value. In languages with FIXED stress (Hyman 1977), the primary stressed syllable has a constant positional relationship to the boundary of a larger domain. Predictable lexical stress can be of two types: *morphological* or *demarcative*. In the former, lexical stress can be determined on the basis of the morphological structure of the word. Such is the case in Russian where stress is assigned with respect to

the stem. In demarcative stress languages, lexical stress can be predicted on the basis of word boundaries. Thus, for example, in Czech, Polish, and French, the stressed syllable is in initial, penultimate, and final position, respectively.

iv. *Morpheme/lexical domain*

Morphemes are strongly constrained with respect to their position within words. Prefixes and suffixes are generally fixed with respect to the stem and to the endpoints of words.

v. *Lexical/lexical domain*

The sequential nature of speech constrains the order of one word with respect to the next. This, of course, is the type of distributional information upon which the postlexical segmentation strategy is based.

vi. *Prosodic/lexical domain*

Finally, some prosodic cues in the signal contribute to the boundary detection process. Thus, a movement in the fundamental frequency can signal word boundaries as well as boundaries of larger domains.

Relational segmentation information

i. *Acoustic-phonetic/lexical domain*

Many languages require agreement in some phonetic property such as vowel harmony, nasalization, or tone within a given domain. In Japanese, a specific tonal melody can be continued throughout a word but not across word boundaries. Similarly, features (like nasalization in Finnish) may spread throughout a word but do not cross word boundaries. Consequently, a change in such features may potentially signal a word boundary. However, when two adjacent words both have this feature, then their shared boundary will of course not be marked by a change in the feature.

ii. *Phoneme/lexical domain*

While the strongest constraints of this type hold for phonemes within the syllabic domain (phonotactic constraints), there are also such constraints within the lexical domain. Trubetzkoy (1977) gives the example of Northern Greenlandic in which constraints on consonant sequences within words exclude the sequence 'stops + consonant' word-internally. Consequently, a word boundary can be inferred to separate a /p/ /f/ consonant sequence, which in many languages can occur word internally.

External sandhi processes provide another important example of relational constraints at this level. Word boundaries can trigger phonological processes that delete, add, and substitute segments. These phonological processes serve in identifying boundaries.

iii. *Syllables/lexical domain*

Since two primary stressed syllables are usually not found in the same word, they could indicate the presence of a word boundary separating them. In addition, the relative duration of syllables or segments also constitutes quantitative relational information. Detailed examination of durational variation in Dutch (Nooteboom 1975) demonstrates the existence of systematic patterns. The duration of a syllable will depend upon its position in the higher-order temporal domain (word or phrase) and also upon the number of other syllables in the domain. Listeners can use these rhythmic effects to determine word boundaries (Nakatani and Schaffer 1977).

All the examples of distributional and relational segmentation information given above signal word boundaries rather than any other boundary. However, since word boundaries are generally shared with other boundaries (phoneme, syllable, morpheme, etc.), it is difficult to determine which higher-level domain is actually being marked. While it is probably true, as Hyman (1978) states, that the word is the most important GRAMMATICAL unit for the statement of cross-linguistic distributional constraints and boundary phenomena, other PHONOLOGICALLY defined domains may be even more important. Indeed, many constraints (e.g. phonotactic) are better defined with reference to prosodic domains (e.g. syllabic) than to the grammatical word.

Research in prosodic phonology (Selkirk 1980; Nespor and Vogel 1982; Vogel 1985) has identified prosodic domains and the way in which they condition phonological processes. Selkirk (1980) distinguishes three types of prosodic domain rules depending upon where with respect to these

domains the rules can apply. One type of rule that is particularly useful for segmentation is the 'domain limit rule' which applies only at the ends of a prosodic domain. Vogel (1985) lists examples of attested prosodic rules for each of the three rule types for six prosodic domains: the syllable, foot, phonological word, phonological phrase, intonational phrase, and phonological utterance. On the basis of her inventory of attested rules, she suggests that there are constraints on the size and number of prosodic domains at which these types of prosodic rules can apply. Similarly, Church (1983) claims that allophonic and phonological processes tend to share the same environment or domain. He illustrates this fact with a large taxonomy of allophonic processes in English defined by their position within the syllable and foot domains. If it is correct that both distributional and relational information tend to signal the same domains, then these domains are obvious candidates for segmentation units.

Cross-linguistic differences in the importance or strength of domains should be reflected by corresponding differences in the segmentation process for these languages. Indeed, the segmentation units for any languages should be the ones that are best marked. Thus, the strength of domains cross-linguistically constitutes a valuable *diagnostic feature* for determining segmentation units in a given language.

Trubetzkoy (1977) claims that languages differ in terms of which higher-level grammatical domain is best marked. He contrasts languages in which morpheme boundaries are most clearly marked with those having more clearly marked word boundaries. He cites German as an example of the former type. All word boundary cues are also morpheme cues, but not the inverse; some cues mark morpheme and not word boundaries. Trubetzkoy calls attention to a second class of languages such as Finnish in which numerous cues signal word and not morpheme boundaries. Shibatani (1973) contrasts distributional constraints for morphemes and words in Japanese. Words constrain phonemes that can occur in final position more than morphemes do. Furthermore, when English words are borrowed into Japanese, word-distributional constraints win over the morpheme ones. Thus, there appear to be differences across languages as to which grammatical unit, the morpheme or the word, is best marked.

Impact of segmentation information on processing models

In the preceding discussion, it has been argued that distributional and relational segmentation information indicates the boundaries and structure of different domains. These domains can be defined GRAMMATICALLY

or PHONOLOGICALLY, and they appear to vary cross-linguistically. These conclusions have major consequences for the models of lexical access presented earlier.

The cohort model operates in a sequential and deterministic fashion. Listeners are assumed to access words one at a time. To do so, they use the beginning of each word to set up a single cohort for that word. In the case of faulty lexical segmentation, an incorrect cohort which does not include the intended word is established. Thus, lexical access cannot take place and lexical segmentation of the following word is impossible. Since correct lexical segmentation is critical in this model, and since the postlexical strategy is not sufficiently reliable to ensure the perfect segmentation required by this model, other sources of information must contribute to the segmentation decisions.

However, the cohort model provides no mechanism for exploiting the type of segmentation information just presented, and thus it needs to be modified. In its present form, it requires that all segmentation information be correct and reliable in marking ONLY word boundaries or the boundaries of the internally stored lexical units. As was just pointed out, any segmentation information for other domains not coterminous with the word would lead to a breakdown in lexical processing. This restriction on the segmentation information is problematic; the distributional and relational information for phonological domains is essentially excluded. For example, even if the syllable were the best-marked domain in a language, it could not serve as a segmentation unit if it straddled word boundaries.

There is an alternative solution in which segmentation information need not be totally reliable and in which a mismatch between segmentation units and internally stored lexical representations does not have disastrous consequences on lexical processing. It is possible to imagine models which allow several segmentations for the same word to be tested with respect to the lexicon. Accordingly, a number of potential word beginnings can be submitted to the lexicon to determine whether they match lexical entries or not. If the cohort based on one segmentation quickly loses all of its members, this segmentation would be rejected and another tried. This approach involves testing only the MOST LIKELY lexical segmentations indicated by the different sources of segmentation information. In other words, well-marked boundaries of prosodic domains such as the syllable or the metrical foot could be tested against the lexicon (see Norris and Cutler, this issue, for a similar proposal).

Although this approach does not require the detected boundaries to match those of the lexical units completely, a close correspondence between them is desirable to minimize the number of incorrect segmenta-

tions to be tested.⁶ The amount of correspondence between the boundaries of the segmentation unit and the internally represented unit is an important factor. There appear to be systematic differences across languages in this correspondence. In many languages syllables and morphemes are not coterminous (i.e. in English, 'farmer', syllable: far.mer, morpheme: farm+er; or in French, 'vivotions', syllable: vi.vo.tions, morpheme: viv+ot+i+ons, where '.' and '+' are syllable and morpheme boundaries, respectively). Similarly, syllable and word boundaries do not always correspond. In some languages like French and English, syllables straddle word boundaries, and in others like Russian, they do not (Garde 1968). The relationship between the boundaries of different domains (coterminous or not) should be examined for different languages since this relationship may serve as another diagnostic for predicting segmentation units. Boundaries of potential segmentation units should generally be aligned with or should have some constant positional relationship to the lexical unit.⁷

There are unfortunately far too few models or experimental tests of segmentation processes to evaluate the suggestion made here. There are only a few models developed in English that have begun to incorporate such segmentation information. For instance, Nakatani (1980) proposed a model which exploits prosodic and rhythmic information in conjunction with allophonic distributional cues. Prosodic information and rhythmic cues indicate to the listener approximately where a boundary is likely to occur. The precise location of the word boundary is signaled by the allophonic cues. These different information sources are monitored simultaneously and continuously to detect these boundaries. An important step toward more systematic exploitation of segmentation information can be found in the finite state parser developed by Church (1983). This parser differs from most other models in that it has an initial parsing stage which segments the speech input into two different domains, the syllable and the metrical foot. Using constraints that the syllable imposes on the distribution of allophones and phonemes (phonotactic constraints), the parser greatly reduces the number of competing syllabic analyses compatible with a given utterance. Nonetheless, some unresolved ambiguity about the correct syllabic segmentation persists despite the effect of these constraints. These ambiguities are resolved by matching the competing syllabic analyses with the lexicon to determine where the words and their boundaries actually lie. Consequently, lexical segmentation and lexical access in this model do not always proceed in a continuous word-by-word fashion but are delayed until after the identification of later words. This model does not rely on the postlexical segmentation information leaving unresolved the fundamental question of whether the integra-

tion of this latter information with that derived from distributional and relational constraints would allow word-by-word recognition.

The models presented are representative of a trend toward the construction of parsers which exploit phonological and prosodic regularities. As more such parsers are developed for different languages, cross-linguistic differences in lexical segmentation and access that depend upon the specific structural properties of the language will begin to emerge. Since languages differ in their phonological, prosodic, and morphological structure, it is reasonable to expect a corresponding difference in lexical segmentation.

Conclusion

This paper has raised the question of how lexical processing and, in particular, lexical segmentation vary across languages. The effectiveness of the postlexical strategy, a potential 'universal' of lexical segmentation, was examined cross-linguistically. According to this strategy, listeners recognize successive words and gain access to their full phonological descriptions prior to hearing them in their entirety. Listeners can thus use this phonological code to predict the end of each word and the onset of the following one. It has been argued that this strategy alone cannot account for every lexical segmentation decision in any language.

The listener must use other types of segmentation information. Two types of segmentation information, distributional and relational, were illustrated with examples from various languages. To exploit either of these information types, a listener must relate events in the signal to specific positions within larger domains. This is possible since lower-level events (phonetic feature, phoneme, etc.) are constrained with respect to where they can occur within higher-level (syllable, morpheme, word) domains. By using knowledge of these constraints the listener can identify the boundaries and internal structure of these domains and segment the continuous speech input.

The contribution of such information and that of the postlexical strategy to lexical segmentation still need to be evaluated carefully. The central role of the postlexical strategy in psycholinguistic thinking is understandable, given the heterogeneity of segmentation information and the lack of processing mechanisms capable of exploiting it. This simple postlexical approach avoids confronting the complicated issues of how relational and distributional information is incorporated into processing models. Nevertheless, since the postlexical approach alone cannot explain

the listener's ability to segment speech, it is important to determine how this segmentation information is combined with that coming from other sources.

Received 12 November 1984

Max Planck Institute, Nijmegen

Revised version received

25 September 1985

Notes

1. An earlier version of this paper was presented at the European Psycholinguistics Association workshop on 'Cross-linguistic studies of morphophonological processing' held in Paris in June, 1984. The contributions from other participants at the workshop are gratefully acknowledged. The author also wishes to thank B. J. Wenk and W. D. Marslen-Wilson for their helpful comments on earlier versions of this paper. Correspondence address: Max Planck Institute, Wundtlaan 1, 6525 XD Nijmegen, The Netherlands.
2. Models developed for auditory word recognition include the logogen model (Morton 1979), the cohort model (Marslen-Wilson and Welsh 1978; Marslen-Wilson and Tyler 1980; Marslen-Wilson 1984, 1985) and the interactive activation model (Elman and McClelland 1984).
3. Most studies of auditory word recognition have dealt with only a few European languages. In particular, English, French, and Dutch have received the most attention.
4. Although the authors of the cohort model have asserted their neutrality with respect to the representation computed from the sensory input, they have generally illustrated their model using phonemic representations.
5. Hyman (1978) also discusses distributional constraints. Trubetzkoy (1977) presents a similar type of segmentation information (*Grenzsignale*) he calls *Einzelsignale*.
6. This view when taken to its extreme eliminates lexical segmentation entirely. Each successive output of the acoustic-phonetic processor is allowed to contact the lexicon and activate a new cohort of words. This position is taken in the TRACE model (Elman and McClelland 1984) in which there is no lexical segmentation. In their interactive activation model, the acoustic input first activates distinctive feature nodes which in turn activate phoneme nodes. The recognition of an individual word takes place when it has been activated to some threshold value. It is only after the word has been recognized that its boundaries can be identified. Thus, there is activation (or classification) of discrete units without prior segmentation of the input.
7. Another possibility should not be neglected. It is conceivable that the internally stored lexical representations themselves are adapted to the phonological structure of the language. As a consequence, the lexical entries could be defined in phonological terms; the mapping could be onto units like the phonological word which in fact can be made up of up to two grammatical words (an open-class entry with its closed-class modifier).

References

- Carlson, R., Elenius, K., Granstrom, B., and Hunnicutt, S. (1985). Phonetic and orthographic properties of the basic vocabulary of five European languages. Paper presented at the Franco-Suedois Symposium, Grenoble, April 20–22.
- Church, K. W. (1983). *Phrase-Structure Parsing: A Method for Taking Advantage of Allophonic Constraints*. Bloomington: Indiana University Linguistics Club.
- Cole, R. A., and Jakimik, J. (1980). A model of speech perception. In R. A. Cole (ed.), *Perception and Production of Fluent Speech*. Hillsdale, N.J.: Erlbaum.
- Elman, J. L., and McClelland, J. L. (1984). Speech as a cognitive process: the interactive activation model. In N. Lass (ed.), *Speech and Language*, vol. 10. New York: Academic Press.
- 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.
- Garde, P. (1968). *L'Accent*. Paris: Presses Universitaires de France.
- Grosjean, F., and Gee, J. (1984). Prosodic structure and spoken word recognition. Working Paper, Northeastern University.
- Hyman, L. M. (1977). On the nature of linguistic stress. In L. M. Hyman (ed.), *Studies in Stress and Accent*. Southern California Occasional Papers in Linguistics 4.
- (1978). Word demarcation. In J. H. Greenberg (ed.), *Universals of Human Language*, vol. 2, 443–470. Stanford, Calif.: Stanford University Press.
- Klatt, D. H. (1980). Speech perception: a model of acoustic-phonetic analysis and lexical access. In R. A. Cole (ed.), *Perception and Production of Fluent Speech*. Hillsdale, N.J.: Erlbaum.
- Lamel, L. F. (1984). The use of structural constraints to determine word boundaries. MIT Speech Communication Group Working Papers IV, 65–72.
- Luce, P. A. (1984). A computational analysis of optimal discrimination points in auditory word recognition. *Research on Speech Perception Progress Report 10*, Indiana University.
- Marcus, S. M., and Frauenfelder, U. H. (i.p.). Word recognition — uniqueness or deviation? A theoretical note. *Language and Cognitive Processes*.
- Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition. In H. Bouma and D. G. Bouwhuis (eds.), *Attention and Performance: Control of Language Processes*. Hillsdale, N.J.: Erlbaum.
- (i.p.). Parallel processing in word recognition. *Cognition*.
- , and Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition* 8, 1–71.
- , and Tyler, L. K. (1981). Central processes in speech understanding. *Philosophical Transactions of the Royal Society of London* B295, 317–332.
- , and Welsh, A. (1978). Processing interactions and lexical access during word-recognition in continuous speech. *Cognitive Psychology* 10, 29–63.
- Massaro, D. W., and Cohen, M. M. (1983). Categorical or continuous speech perception: a new test. *Speech Communication* 2, 15–35.
- Mehler, J., Segui, J., and Carey, P. (1978). Tails of words: monitoring ambiguity. *Journal of Verbal Learning and Verbal Behavior* 17, 29–35.
- Morton, J. (1979). Word recognition. In J. Morton and J. C. Marshall (eds.), *Psycholinguistics 2: Structure and Processes*. Cambridge, Mass.: MIT Press.
- Nakatani, L. H. (1980). Fission and fusion processes in speech perception. Unpublished manuscript, Bell Labs, Murray Hill, N.J.
- , and Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America* 62 (3), 714–719.

- , and Schaffer, J. A. (1977). Hearing 'words' without words: prosodic cues for word perception. *Journal of the Acoustical Society of America* 63 (1), 234–245.
- Nespor, M., and Vogel, I. (1982). Prosodic domains of external sandhi rules. In H. van der Hulst and N. Smith (eds.), *The Structure of Phonological Representations*. Dordrecht: Foris.
- Nooteboom, S. G. (1975). On the auditory representation of syllable nucleus durations. In G. Fant and M. A. A. Tatham (eds.), *Auditory Analysis and Perception of Speech*, 413–430. London: Academic Press.
- Pisoni, D. B., and Tash, J. (1974). Reaction times to comparisons within and across phonetic categories, *Perception and Psychophysics* 15 (2), 285–290.
- Selkirk, E. O. (1980). Prosodic domains in phonology: Sanskrit revisited. In M. Aronoff and M.-L. Kean (eds.), *Juncture*, 107–129. Saratoga, Calif.: Anma Libri.
- Shibatani, M. (1973). The role of surface phonetic constraints in generative phonology. *Language* 49, 87–106.
- Studdert-Kennedy, M. (1980). Speech perception. *Language and Speech* 23 (1), 45–66.
- Taft, L. A. (1984). Prosodic constraints and lexical parsing strategies. Unpublished Ph.D. thesis, University of Massachusetts at Amherst.
- Tanenhaus, M. K., and Lucas, M. M. (1985). Context effects in lexical processing. Unpublished manuscript.
- Trubetzkoy, N. S. (1977). *Grundzüge der Phonologie*. Gottingen: Vandenhoeck and Ruprecht.
- Vogel, I. (1985). On constraining prosodic rules. In H. van der Hulst and N. Smith (eds.), *Advances in Non-Linear Phonology*. Dordrecht: Foris.
- Zue, V. W., and Schwartz, R. M. (1980). Acoustic processing and phonetic analysis. In W. A. Lea (ed.), *Trends in Speech Processing*. Englewood Cliffs, N.J.: Prentice-Hall.

