How long is the sentence? Prediction and prosody in the on-line processing of language*

FRANÇOIS GROSJEAN

Abstract

The gating paradigm (Grosjean 1980) was used to determine whether subjects who are listening to the 'potentially last word' of a sentence (in this case, a noun before an optional prepositional phrase) can indicate whether the sentence is over or not, and if it is not over, how much longer it will last. Sentences that contained endings ranging in length from zero to nine words were gated on the object noun and presented to subjects who had to choose which of four sentences was being presented or press a key at a point in time when they felt the sentence would have ended had it been presented in full. Results showed that, basing themselves solely on prosodic cues, subjects were surprisingly accurate at predicting the length of the upcoming endings. An acoustic analysis of the test sentences showed a strong relationship between measures of fundamental frequency, amplitude, and duration and the experimental data. These findings are discussed in terms of the predictive and interpretative roles of prosody during the on-line processing of language.

Much of current research in psycholinguistics is aimed at understanding the processes by which language is perceived in real time. Researchers are studying the time course of language perception and understanding; that is, how the listener goes from the acoustic signal to an interpretative representation of the message. Part of this research involves understanding how the processing of the acoustic signal interacts with the discourse context and the rules of the language as well as with pragmatic and environmental factors. A number of assumptions underlie this research. One is that language processing is obligatory and takes place on-line or in real time (Marslen-Wilson and Tyler 1980). Another is that the listener will use any and all information that may facilitate this processing. A third is that the various sources of information will interact — at some level at least — during processing. A fourth premise — and one that

Linguistics 21 (1983), 501–529 0024–3949/83/0021–0501 $2.00

has received less attention — is that the listener will use past and current information not only to process the utterance up to the point reached by the speaker but also to predict upcoming information. It is this premise that we will attempt to develop in this paper.

The phenomenon of prediction during language processing has been referred to in a number of ways: anticipation, expectation, assumption, predictive tracking, feedforward, look ahead, forecasting. Whatever the label, prediction is a phenomenon which becomes apparent when the listener finishes a sentence for the speaker, when he or she replies before the speaker is finished, or when the listener tells the speaker, 'I knew exactly what you were going to say.' It also becomes apparent when a person does not reply even though the sentence is finished both syntactically and semantically; this happens when the prosodic information tells the listener that more is to come. In fact, prediction is pervasive throughout the give and take and the turn-taking of conversations: people rarely interrupt one another, but they also rarely leave long gaps between exchanges (Brazil 1981).

Prediction is facilitated by the structure and the rules of the language. In English, at the level of discourse, old information usually precedes new information, the topic usually precedes the focus, definite items often come before nondefinite items, and human reference usually precedes nonhuman reference (Comrie 1981). In some languages, these phenomena are actual linguistic rules. At the level of syntax, English contains rules that can help predict — at least in part — the next word, phrase, or sentence. Subject noun phrases are usually followed by verb phrases; certain transitive verbs are followed by adverbial or prepositional phrases or new sentences; other transitive verbs take only a NP object. Two-clause sentences are often marked at the beginning in such a way as to indicate the upcoming construction. Thus, function words usually mark coordinate, subordinate, and relative clauses (Clark and Clark 1977). At the level of the word, the sequence of sounds will often allow the word to be distinguished from other words long before its ending and context will narrow down the set of word candidates and hence accelerate its recognition (Marslen-Wilson and Welsh 1978). Finally, at the levels below the word, coarticulation phenomena as well as phonological and morphological rules often give cues as to the next phoneme or next morpheme. To these structural aspects of the language can be added conversational rules, pragmatics, and knowledge of the world, as well as customs and traditions. All this indicates that the rule-governed aspects of language as well as the psychosocial environment in which language is used favor prediction during on-line processing.

The question now is, in what ways can prediction be helpful to the listener? Three answers can be proposed. First, prediction can reduce the set of possibilities and therefore can focus the attention of the listener. This in turn makes processing more efficient and even accelerates it, be it during speech perception, during word recognition, or during syntactic parsing, for example. Second, prediction can help demarcate or set up a domain of processing. It can, for example, mark a word boundary or a phrase or sentence boundary, so that the listener is ready for the next word, phrase, or sentence. And third, prediction can give the listener time for other activities such as integrating the information, packaging it off, preparing a response, and so on.

A number of points need to be kept in mind when talking about prediction. The first is that it is rarely perfect; prediction is probabilistic by definition. The listener needs to take this into account if or when prediction takes place. The second point is that prediction need not take place during on-line processing. One can easily imagine a processing system that only uses past and present information and never uses this information to look ahead. And the third, and related, point is that there is as yet very little direct evidence for the use of prediction in processing. This is mainly because few studies have attempted to examine this topic. Information and communication theory led researchers in the 1950s to study the statistics of language as well as stochastic chains and readability formulas, but the impact of transformational-generative grammar on psycholinguistics in the 1960s practically put an end to this work. The concept of prediction was brought back in part in the early 1970s with the work on processing strategies (Bever 1970; Kimball 1973; Clark and Clark 1977). Some of these strategies do have a very strong look-ahead component, such as strategy 2 proposed by Clark and Clark (1977: 61): 'After identifying the beginning of a constituent, look for content words appropriate to that type of constituent', or strategy 4 (1977: 64): 'After encountering a verb, look for the number and kind of arguments appropriate to that verb.' On the whole, however, few current models of perception and comprehension integrate prediction as an indispensable and ever-present aspect of language processing.

One important source of information during language processing is prosody; that is, the intonation, stress, rhythm, loudness, and rate of the utterance. Prosodic information can be used either to process the signal up to the point reached by the speaker (and hence by the listener) or to predict what is to come beyond that point. As an example of the first alternative, we can mention the role of prosody in sentence parsing. It is a well-known fact that prosody is sometimes the ONLY cue that is available to the parser. This is the case, for example, when two sentences are separated by an adverbial or prepositional phrase which can be part of either the first or the second sentence. For example,

[The other day I saw John really anxious]$_S$ [I said 'Hi' to him]$_S$
[The other day I saw John]$_S$ [Really anxious, I said 'Hi' to him]$_S$

In this example, apart from some possible coarticulatory information, only the prosody can help the listener parse the sentence correctly; no other information, be it syntactic or semantic, can do so. Another example involves ambiguous phrases such as

What is [two plus three] times four?
What is two plus [three times four]?

Here again only the prosody will lead the listener to propose either 20 or 14 as the answer. A series of studies (Lehiste 1973; Streeter 1978; Scott 1982) has shown that sentences of this type are indeed disambiguated by prosody and that certain variables such as duration (or rate) and fundamental frequency play a greater role than others (such as amplitude) in this disambiguation. Despite this work, however, parsing models like those of Foder et al. (1974), Kimball (1973), or Frazier and Fodor (1978) do not include a prosody component either to help define and delineate the basic parsing units or to help the second-stage parsers in combining these units. And general processing models such as the one proposed by Marslen-Wilson and Tyler (1980) rarely mention prosody — either as an aid to word recognition and syntactic parsing or as a carrier of rich semantic information.

Prosody can also help the listener predict later events. Two researchers — James Martin and Anne Cutler — are most closely identified with this predictive aspect of prosody. Martin (1972; Shields et al. 1974; Meltzer et al. 1976) proposes that speech is rhythmically structured so that the location of later accents in the sentence can be predicted from the hierarchical organization of the earlier accents. The relationship between stressed syllables and timing is such that aspects of the signal become predictable in time. The listener 'locks into' the stress pattern and this allows him or her to anticipate upcoming stressed syllables in the speech stream as well as the end of the speech sequence. Evidence for this position comes from a number of studies conducted by Martin and his colleagues in which subjects monitored phonemes in nonsense words as well as in words displaced in time (Shields et al. 1974; Meltzer et al. 1976).

Cutler (1976; Cutler and Darwin 1981; Cutler and Fodor 1979; Cutler and Foss 1977) also believes that the prediction of upcoming accents is an integral part of sentence processing. She agrees with Martin that one role for prosody is to direct the attention of the listener to the main accents of the sentence. Cutler adds that the listener needs to find the focus of the sentence because it carries the new information, and he or she will do so by cueing in on the prosodic information. Evidence for this position comes from a study in which Cutler (1976) showed that the reaction times to the first phoneme of a prosodically neutral word were faster when the word was spliced into a context predicting high stress than when spliced into a low-stress context.

In the present study we will investigate further the information that is carried by the prosody of an utterance concerning later events in that same utterance. A brief anecdotal observation will help set the stage. Any listener of radio news programs will have noticed from time to time that speakers who are interviewed are sometimes cut off midway through their utterance by the editor preparing the interview for broadcasting. A closer examination of the splicing points reveals that whereas the main syntactic and semantic boundaries were respected, the prosodic breaks were not. That is, the spliced utterance makes perfect sense but prosodic cues such as intonation, rate, and rhythm tell the listener that the utterance is not over. In fact, the listener sometimes has the impression that the speaker said much more than was actually given. It is this very phenomenon — the information carried by the prosody about the length of the utterance — that we will examine here.

The present study has several aims. The first is to determine whether subjects who are listening to the 'potentially last word' of a sentence can indicate whether the sentence is over or not, and if it is not over, how much longer it will last. Can this be done solely with prosodic information? Shields et al. (1974) expect this to be the case. They write, 'once early accents are heard, the location of later accents (and perhaps the end of the pattern as well) can be predicted in time' (1974: 251, col. 1). Breckenridge (1978) concurs with this position and writes concerning intonation,

> Since the slope of the declination is adjusted so that a fixed terminal point is not reached until the end of the intonation group, it might even the the case that speakers can extrapolate a declination line in order to anticipate how much is still to come (1978: 45).

In the present study we will use sentences that take optional prepositional phrases and will vary the length of these phrases. We will splice these out and will ask subjects at the end of the segments that remain to indicate how much longer the sentences continue. Because syntax and semantics have little to say about optional prepositional phrases, subjects will have to reply solely on prosodic cues when making their judgments. Thus we will be able to determine whether the prosody at a point in time in the sentence carries information about the length of the remaining part of the sentence. If this is the case, it will be an interesting finding, as this type of prosodic information does not only inform the listener about the length of the upcoming unit, but can also help him or her delineate the end of the present sentence and hence the beginning of the new one. This, in turn, will facilitate parsing and end-of-sentence packaging and will also help the listener prepare a response if he or she has decided to intervene.

A second aim of the study is to determine at what point during the 'potentially last word' the prosodic information becomes available to the listener and/or can be used. Is it throughout the word or only at the end of the word? To answer this question we will 'gate' the potentially last word, that is the word just before the prepositional phrase, so that at the first presentation the listener will not hear any of the word. At the second presentation he or she will hear the first 50 msec of the word, at the third presentation the first 100 msecs, and so on, until the whole word has been presented. (For information concerning the 'gating' approach, see Grosjean 1980; for earlier versions of the paradigm, see Pollack and Pickett 1963; Ohman 1966).

A third aim of the study is to show that the results obtained are not response-specific. To do this we will use a multiple-choice response task as well as an open-ended response task. We hope that the responses obtained with the two different tasks will be as similar as possible so as to allow us to eliminate as much of the task component from the results as possible.

A fourth aim is to determine how well the results obtained can be predicted by an acoustic analysis of the sentence up to and including the potentially last word of the sentence. We will correlate the responses obtained from our subjects with measures of fundamental frequency, duration, and amplitude and observe whether there is any relationship between the psychological measures and the acoustic measures. We should note that any strong relationship is not a necessity. It could well be that listeners integrate the acoustic information in a complex way (as is done for the perception of stress and rhythm, for example) and that this is not reflected by any one of our 'simple' acoustic measures, or simply that we have not chosen the appropriate measures.

A final aim of the study is to examine how the gating approach used here can be extended to other questions involving prosodic information in the speech stream. We believe that this approach can help determine the amount of prosodic information that is available to the listener at a particular point in the utterance concerning various aspects of the remaining, and as yet unheard, part of the utterance. In addition, it can determine whether this information can be used by the listener when making a response in a laboratory situation. It does not, however, allow one to show that listeners do in fact use this information during on-line processing, and if they do, when and how. Only the use of on-line experimental paradigms will enable us to answer these questions. The present study is only a first step in this direction.

Experiment 1

Method

Subjects. Thirty-two students, with no reported speech or hearing defects, served individually in sessions lasting 30 minutes.

Materials. Thirty-two sentence exemplars were used in the experiment (see the Appendix). Each exemplar belonged to one of four types of sentences. The first type (the zero-word-ending type) was a six-word simple declarative sentence. It started with a sentential adverb and ended on an object noun. Its verb favored but did not mandate a prepositional phrase and was therefore subcategorized as ⟨–NP (PP)⟩. Three of the eight examplars that made up this type of sentence are

Yesterday my sister made a cake.
Earlier my sister took a dip.
Yesterday the person found a bike.

The second type of sentence (the three-word-ending type) was identical to the first except that it continued with a three-word prepositional phrase. Appropriate phrases were added to each of the Type 1 exemplars, as in

Yesterday my sister made a cake for the fair.
Earlier my sister took a dip in the pool.
Yesterday the person found a bike on the rack.

The third type of sentence (the six-word-ending type) was again identical to the first except that it continued with a six-word prepositional phrase. Here a three-word phrase was embedded in the prepositional phrase of the three-word-ending type. For example,

Yesterday my sister made a cake for the fair at the school.
Earlier my sister took a dip in the pool at the club.
Yesterday the person found a bike on the rack in the shed.

Finally, the fourth type of sentence (the nine-word-ending type) continued with a nine-word prepositional phrase; here again the additional three-word phrase was embedded in the preceding prepositional phrase. This gave

Yesterday my sister made a cake for the fair at the school on the hill.
Earlier my sister took a dip in the pool at the club on the hill.
Yesterday the person found a bike on the rack in the shed of the yard.

The 32 sentence exemplars were recorded by a female speaker who was asked to read each exemplar in one breath group, with no breaks such as pauses in the prosodic pattern. The exemplars were then spliced with the aid of a PDP 9 computer. Each exemplar was digitized at a sampling rate of 10 kHz and manipulated by means of a computer-controlled cursor. First the onset of the exemplar and the beginning and end of the 'potentially last word' were marked off. This is the last word of Type 1 exemplars and the word before the prepositional phrase in Types 2, 3, and 4 exemplars. Thus, in the examples above, it is 'cake', 'dip', and 'bike'. As this word always started and ended with a stop consonant, it was relatively easy to mark off its onset and offset. The onset was defined as the beginning of the release burst of the word's initial plosive and the offset as the end of the release burst of the word's final plosive.

Markers were placed every 50 msec of the stimulus word; when the word's duration was not an exact multiple of 50, the last gate (or segment) length was the difference between the second-to-last gate time and the duration of the word. Each exemplar was now output in presentations of increasing duration. Thus, the first presentation corresponded to the exemplar up to the beginning of the stimulus word but did not contain any of the burst of the initial stop. The second presentation corresponded to the exemplar followed by 50 msecs of the stimulus word; the third presentation included 100 msecs of the word, and so on, until at the nth presentation, the exemplar and the whole of the last word were presented. Each sentence exemplar was therefore represented by a presentation set where the first presentation contained none of the potentially last word of the sentence and the last presentation contained the whole of that word. (No set, therefore, ever presented the prepositional phrases that ended three of the four sentence types.) All 32 presentation sets were prepared in this way, eight for each of the four types of sentences.

Four experimental tapes were prepared. Each tape contained eight presentation sets, two from each of the eight exemplars of each of the four sentence types. The sets were presented in random order and the same sentence beginning (for example, 'Yesterday my sister made a ...') never appeared twice on the same tape.

Procedure. Four groups of eight subjects were run, one group for each of the experimental tapes. Subjects were presented with eight answer sheets, one for each presentation set. At the top of each sheet were the exemplars of the four sentence types that had the same beginning. Each exemplar was preceded by a letter ranging from 'a' to 'd'. For example,

a. Yesterday my sister made a cake.
b. Yesterday my sister made a cake for the fair.
c. Yesterday my sister made a cake for the fair at the school.
d. Yesterday my sister made a cake for the fair at the school on the hill.

Below these exemplars were an array of numbered lines containing the letters 'a, b, c, d' and a 1–10 confidence rating scale where 1 was labeled 'very unsure' and 10 'very sure'. There were 15 such lines in all.

Subjects were asked to listen to each presentation set and to indicate after each individual presentation whether the sentence segment that had been presented came from sentence a, b, c, or d. To do this they were asked to circle the appropriate letter on the answer sheet and to indicate how sure they felt about their guess by circling a number on the unsure–sure scale. Each presentation set was announced by a number and subjects were asked to turn to the next answer sheet at the beginning of each new set. Subjects were informed that each presentation set was based on just one sentence type ending, but that they were free to change their guesses as they progressed through a set. No previous guess could be changed, however.

Data analysis. Because the eight potentially last words varied in duration within and between sentence types, all gate durations were transformed to a percent of the way through the word (% WT). To do this, each gate duration for a particular word was divided by the length of the word and the product was multiplied by 100.

Five gates were chosen for data analysis. They were those at or closest to the 0, 25, 50, 75, and 100% WT marks. The letter circled by each subject at these gates was transformed to a number: 'a' corresponded to zero more words estimated to the end of the sentence, 'b' to three more words estimated, 'c' and 'd' to six and nine more words estimated. The confidence ratings given by subjects at each of these gates were also recorded. Results were averaged both over sentence exemplars and over subjects, and two separate analyses of variance were run on the number of words estimated and the confidence ratings. In the first, sentence type (*ending*) and gate duration (*gate*) were fixed effects and sentence exemplar (*sentence*) was a random effect, and in the second the fixed effects were again *ending* and *gate* but *subject* was the random effect. This allowed the subsequent calculation of a min F′ statistic for *ending*, *gate*, and the *ending* × *gate* interaction, both for the number of words estimated to the end of the sentence and for the confidence ratings.

Results and discussion

Figure 1 presents the estimated number of additional words to the end of the sentence as a function of the percent of the way through the word for each of the four sentence types. A visual examination of the figure reveals that at the beginning of the word (the 0% gate), subjects cannot differentiate between the four sentence types but that as they progress through the word they become more and more proficient at doing so, especially for the sentences with zero-, three-, and six-word endings. At the last gate (100% of the word) the estimated values are surprisingly close to the real values: sentences that stop on the word (the zero-word ending) are estimated to continue for 1.03 words on average, sentences with three-word endings are estimated to continue for 3.8 words, and those with six-word endings are estimated to continue for exactly 6.0 words. Only the nine-word-ending sentences are off the mark; subjects estimate that they continue for 5.1 words on average.

A min F′ statistic based on the two analyses of variance that were run confirms this visual pattern. First, a main effect was found for *ending* [min $F'(3, 54) = 5.52$, $p < 0.01$], indicating that subjects are capable of differentiating between the length of various sentence endings before these endings are actually heard. An *a posteriori* test (Tukey HSD, Kirk 1967,

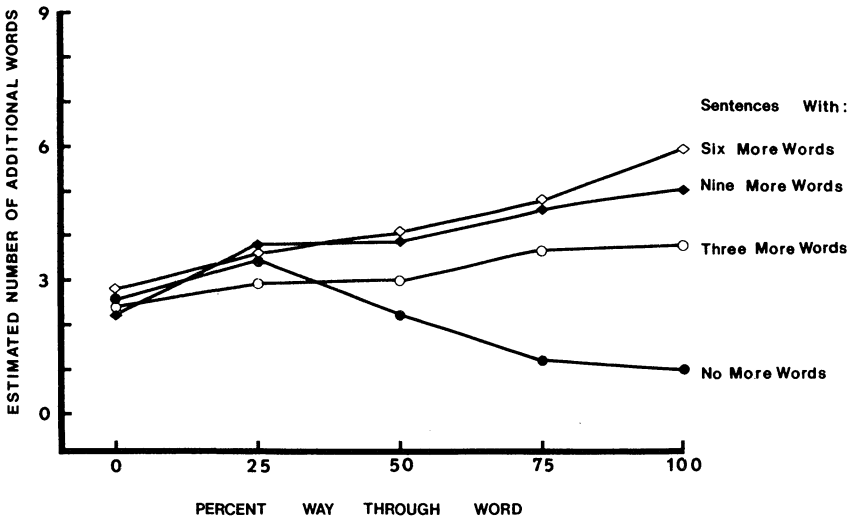

Figure 1. *The estimated number of additional words to the end of the sentence as a function of the percent of the way through the word for each of the four sentence types. Each point is the mean of 64 observations, two by each of 32 subjects*

$p<0.05$) on the ANOVA with subjects as the random factor showed that the global mean for the zero-word ending (2.1 words) was significantly different from that for the three-word ending (3.2 words) and in turn that the mean for the three-word ending was significantly different from that obtained for the six-word ending (4.3 words). The mean for the nine-word ending (4.0) was not different, however, from that for the six-word ending.

Second, a main effect was obtained for *gate* [min F' $(4, 77)=4.47$, $p<0.01$], indicating a general increase in the number of words estimated as subjects work their way through the stimulus word. An *a posteriori* test (Tukey HSD, $p<0.05$) reveals that this effect is only due to the sudden increase in the number of words estimated between the 0% gate (2.6 words estimated) and the 25% gate (3.5 words) and not to the differences between later gates.

Third, and more importantly, a significant *ending* × *gate* interaction was found [min F' $(2,247)=7.71$, $p<0.01$], confirming the slow differentiation pattern observed in Figure 1. An *a posteriori* test (Tukey HSD, $p<0.05$) at each test gate also confirms this pattern. No mean is different from any other mean at the 0% and 25% gates. At the 50% gate, the mean of the zero-word ending is not yet different from that of the three-word ending, but it is now different from the mean of the six-word ending. At the 75% gate, all three endings (zero, three, and six words) are significantly different from one another, and this difference grows larger at the 100% gate. As for the nine-word ending, it is significantly different from the three-word ending at the 50% gate and at the 100% gate but is never different from the six-word ending.

During the experiment, subjects not only guessed which sentence was being presented to them, they also rated the confidence they had in their guess. Figure 2 presents the mean confidence rating given by subjects as a function of the percent of the way through the word for each of the four sentence endings. Several points can be made based on the pattern obtained. The first is that confidence ratings increase as subjects hear more of the word: the global mean is 3.2 at the 0% gate and 6.1 at the 100% gate. A min F' statistic confirms this [min F' $(4, 147)=45.51$, $p<0.01$] and a Tukey *post hoc* test ($p<0.05$) reveals that the increase is significant at all gates except between the 0% gate and the 25% gate. The second point is that the confidence ratings for the zero-word-ending sentences increase more rapidly than for the other three types of sentences. This is confirmed not only by an *ending* main effect [min F' $(3, 37)=6.81$, $p<0.01$] but especially by an *ending* × *gate* interaction [min F' $(12, 240)=3.57$, $p<0.01$]. A *post hoc* Tukey test shows that all mean ratings are identical at the 0% gate and that only the zero-word-ending

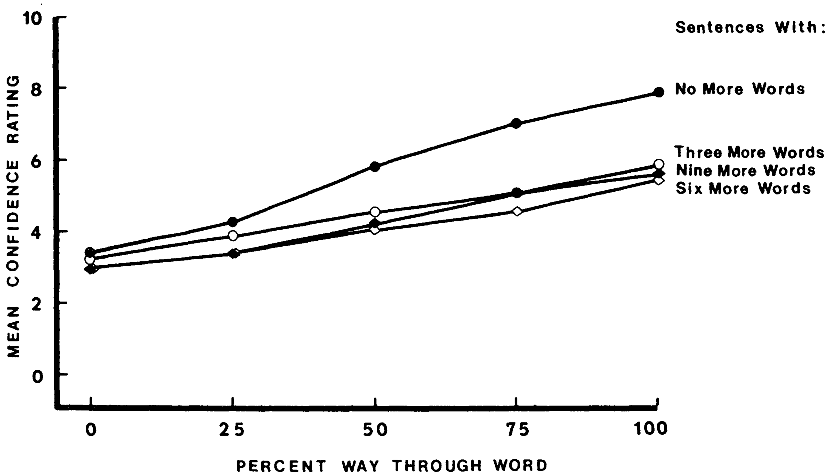

Figure 2. *The confidence rating of ending length estimates as a function of the percent of the way through the word for each of the four sentence types. Each point is the mean of 64 observations, two by each of 32 subjects*

mean is significantly different from the other means at the 100% gate (7.8 for the zero-word ending as compared to 5.8, 5.4, and 5.6 for the three-, six-, and nine-word endings respectively). This pattern signifies two things. First, the rather low and equal confidence ratings obtained for the various sentence endings at the beginning of the stimulus word seem to confirm that there is little prosodic information available at that point concerning the length of the sentence (see Figure 1). Second, and not surprisingly, the rich prosodic information that marks the end of a sentence in the zero-word-ending sentences leads subjects to feel much more confident in their response than does the less rich information linked to sentences that continue for three, six, or nine more words. It is nonetheless interesting that, faced with the difficult task of estimating the length of the sentence, subjects should be able to reach a mid value in the confidence rating scale. This in itself must mean not only that prosodic information is present at that point, as is revealed in Figure 1, but that subjects are overtly conscious of it.

Four main conclusions emerge from the results so far. The first is that subjects can be extraordinarily precise in estimating the length of a sentence when they are only part of the way through the sentence. Not only can they indicate whether a sentence is over or not (a fact that is well known in the literature), but they can also distinguish between a sentence that continues for three more words and one that continues for six more words. This clearly shows that the prosody of a sentence carries information about the length of that sentence, as has been suggested by Shields et al. (1974) and Breckenridge (1978).

The second point is that subjects cannot differentiate sentences with a nine-word ending from those with a six-word ending. This is probably because the former sentences in our study were abnormally long (15 words) and were read in one breath group. As we will see later, the actual prosodic values of these sentences, in terms of fundamental frequency, amplitude, and duration, are situated between those of the three-word and those of the six-word endings, and thus it is not surprising that the estimates fall where they do (at the 100% gate, the mean estimate for the nine-word ending was 5.1 words).

The third point is that subjects cannot distinguish between the various types of sentences at the beginning of the potentially last word of the sentence. This is true not only for the sentences with three-, six-, and nine-word endings but also, and more surprisingly, for the sentences that are actually ending. In other words, subjects are unable at that point to distinguish sentences that will be ending 300 msecs later from those that will be continuing for another 1400 msecs!

And the final point is that a slow and steady differentiation of the sentence types takes place between the beginning and the end of the stimulus word. This differentiation, which leads to surprisingly precise estimations at the end of the word, is also reflected in the confidence ratings. Subjects feel more and more confident about their guesses as they progress through the word.[1]

A number of reasons can be proposed to explain the lack of differentiation at the beginning of the word but the very good differentiation at the end of the word. Two of these are linked to the paradigm itself and will be examined in detail here. The first reason is that subjects may be adopting a 'wait and see' strategy at the first gates of each presentation set. They know that with each presentation they will hear more and more of the word and so they decide to give rather neutral ending estimates to the first gates. They do this even though some prosodic information concerning the endings may be present at these gates. The second reason for the slow differentiation is that subjects are being influenced by the repetitive nature of the paradigm (the early part of the word is repeated over and over again). It could be that subjects get better at differentiating the endings precisely because the gates have a repetitive aspect to them.

In order to determine whether these two explanations were the correct ones, we ran one group of eight new subjects on all 32 0% gates and a second group of eight subjects on all 32 100% gates. Sentence types and exemplars were randomized in each case and subjects were run according to the same procedure as in the main experiment.

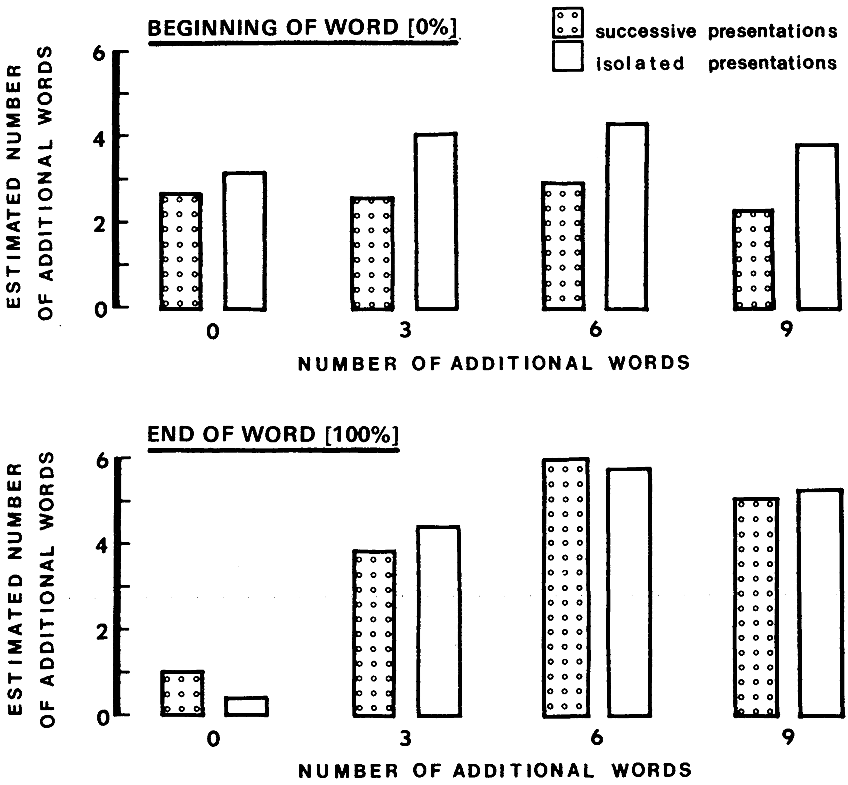

Figure 3. *The estimated number of words to the end of the sentence at the 0% gate (top part) and at the 100% gate (bottom part), as a function of the four sentence types. The filled bars represent the means of the experimental group (64 observations, 2 by each of 32 subjects) and the empty bars those of the control group (32 observations, 4 by each of 8 subjects)*

Figure 3 presents the results obtained from these two control studies. In the top part of the figure the estimated number of additional words at the 0% gate are plotted as a function of the four sentence endings. The results from the main study (successive presentations) are presented next to those of the control study (isolated presentations). Two observations can be made. The first is that the groups differ significantly on the number of words estimated to the end of the sentence. The global mean is 3.9 words for the new group and 2.6 words for the old group [F (1, 14) = 14.14, p < 0.01]. This difference, which is not of direct interest to us here, may be due to a number of factors such as the number of gates per word heard by subjects (one in the isolated presentation versus several in the successive presentation), the number of sentence endings given to subjects (32 as compared to 8), or the subjects themselves. Of greater interest is that no main effect was found either for *ending* [$F(3, 42) = 2.51$, N.S.] or for the *ending* × *group* interaction [$F(3, 42) = 1.47$, N.S.].[2] Thus the lack of differentiation in the original study at the 0% gate does not seem to be due to a strategy of 'wait and see' on the part of the subjects. Rather it would appear that the relevant prosodic information is either not present at the 0% gate or cannot be used by subjects at that particular gate.

The bottom part of Figure 3 presents the estimated number of additional words at the 100% gate as a function of the four sentence endings. As can be seen, the subjects who worked their way through every gate of the words in the successive presentations and those who only heard the last gate of the words in the isolated presentations give practically identical results. There is, as expected, a strong main effect for *ending* [$F(3, 42) = 73.84$, $p < 0.01$] but no effect for *group* [$F(1, 14) < 1$] and no *ending* × *group* interaction [$F(3, 42) = 1.15$, N.S.]. Thus the slow differentiation that takes place as subjects work their way through the word is not due to the repetitive nature of the gating task.

These results, which show that the differentiation pattern is not caused by the paradigm itself, call for other explanations. One of these is that the prosodic information concerning the remaining part of the sentence is available only during stressed syllables and not during unstressed syllables. (It is a well-known fact that it is the stressed syllable of a word that carries most of the prosodic information.) At the early gates, subjects have only the immediate prosodic information carried by the unstressed determiner and little, if any, information carried by the stressed stimulus word. It is only as they progress through the word and reach the stressed vowel that the information becomes available and that their responses can reflect it. Linked to this explanation must be the fact that estimates based on preceding stressed syllables cannot be carried over to unstressed syllables. As a consequence, at each new stressed syllable, subjects have to reestimate the length of the sentence. To test this explanation one will want to gate the test sentences through stressed and unstressed syllables to see if the information available on a stressed syllable is indeed no longer available on an unstressed syllable.

A second possible explanation is that subjects can only estimate the length of a sentence when they are listening to the potentially last word of that sentence. If, on reaching a word, the syntax or semantics tells them that the sentence must continue, then the prosodic information concerning the length of the remaining segment is either not made available to the listener or cannot be accessed by the listener. It is only when a potentially last word is reached that it is critical for the listener to know whether the next word is the last word of the sentence or not. It is at this time, therefore, that the prosodic information is made available or is called into play.[3] Of course, the listener will have to wait for the first stressed syllable of that word to start making correct predictions. To test this explanation one will want to obtain estimates of ending length from subjects who are listening to two different types of stressed words — those that are potentially the last word of the sentence and those that are not.

In sum, the pattern of results in this experiment is clear — no differentiation between the endings at the beginning of the word and accurate differentiation at the end — but the explanation for this pattern will have to await further experimentation.

Experiment 2

Before attempting to determine how well our results could be predicted by an acoustic analysis of the sentences, we wanted to make sure that our findings were not specific to the response task used by the subjects. The multiple-choice task used in the experiment had a number of characteristics which could have biased the subjects in their responses. For example, the correct response was given on the answer sheet along with three incorrect responses. In addition, the four sentence endings were presented and subjects may well have preferred one ending over another for semantic reasons. Also the nine-word endings probably 'read' less well than the three- or six-word endings. Could the same pattern of results be obtained with a response task that did not contain these characteristics? In order to answer this question, and to bring converging evidence to our findings, we ran a second experiment in which subjects were asked to listen to the sentence fragments and to press a key when they thought the sentence would have ended had the sentence been given to them in its entirety.

Method

Subjects. Twenty-four of the 32 subjects who had taken part in experiment 1 were brought back three months later and were run individually in sessions lasting 30 minutes.

Materials. The four tapes of experiment 1 were used. Each tape contained eight sets of sentence fragments.

Procedure. Four groups of six subjects were run, one group for each experimental tape. Subjects were assigned to a tape that they had not heard in the first experiment. In the experimental sessions, subjects listened to the eight sets of sentence fragments and were asked, after each individual presentation, to press a key at the point in time when they thought the sentence would have ended had it been presented in its entirety. Thus, if they felt the sentence ended after the stimulus word, they were to estimate the length of that word and then press the key. On the other hand, if they felt the sentence continued for some time, they were to wait for that amount of time and only then press the key.

The eight sentence fragments as well as the stimulus words (but not the prepositional phrases) were presented to the subjects in writing at the beginning of the experiment and remained in front of them during the session. This was done in order to help subjects concentrate solely on estimating the length of the last part of the sentence rather than being involved in a word-recognition task.

The beginning of each sentence fragment triggered a voice-operated relay which started a clock on a PDP 11–10 computer and the key press stopped the clock. Each of these times was subtracted from the duration of the sentence fragment at that presentation so as to obtain the estimated duration of the sentence ending.

Data analysis. The estimated durations of the sentence endings were analyzed, in each presentation set, at the five gates chosen in Experiment 1 (that is, those gates closest to the 0, 25, 50, 75, and 100% points). Responses were averaged once again over sentence exemplars and over subjects and two separate analyses of variance were run. This allowed for the computation of a min F′ statistic for *ending*, *gate*, and the *ending* × *gate* interaction.

Results and discussion

Figure 4 presents the estimated time to the end of the sentence as a function of the percent of the way through the word for each of the four sentence types. As can be seen, there is a very close similarity between the results obtained with this key-press task and those obtained with the multiple-choice measure (Figure 1). Both sets of results show a lack of differentiation between the various ending types at the early gates and a steady differentation as subjects progress through the word. The statistical similarity of the two sets of results is reflected in the following ways. First, there is a 0.87 correlation between the means at each of the five test gates for each of the four sentence types in the two tasks ($N = 20$). Second, a min F′ statistic on the key-press data shows very similar main effects. A

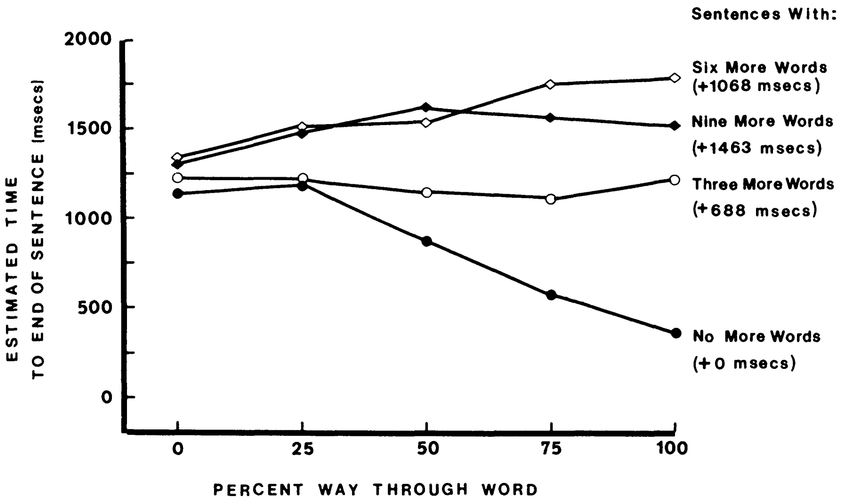

Figure 4. *The estimated time to the end of the sentence as a function of the percent of the way through the word for each of the four sentence types. Each point is the mean of 48 observations, two by each of 24 subjects*

main effect was again found for *ending* [min F′ (3, 30) = 4.86, $p < 0.01$] and a Tukey *post hoc* test ($p < 0.05$) on the ANOVA with subjects as a random factor showed that the global mean for the zero-word ending (837 msecs) was significantly different from that for the three-word ending (1189 msecs), and in turn the mean for the three-word ending was significantly different from the mean obtained for the six-word ending (1587 msecs). The mean for the nine-word ending (1510 msecs) was not different from that for the six-word ending, but it was different from the mean for the three-word ending.

No main effect was found this time for *gate* [min F′ (4, 120) = 0.65, N.S.], but once again a strong main effect was found for the *ending* × *gate* interaction [min F′ (12, 223) = 3.29, $p < 0.01$]. A Tukey *post hoc* test at each test gate confirms the pattern of slow differentiation. At the first two gates (0% and 25%), no mean is different from any other mean. At the 50% gate, the zero-word ending is now different from the six-word ending, and at the 75% and 100% gates, the three endings (zero, three, and six words) are all different from one another. As for the nine-word ending, it is never different from the six-word ending, but it is different from the three-word ending at the 50% and 75% gates. In sum, these results bring converging evidence to the finding that subjects can estimate the length of a sentence when they are presented with only part of that sentence. Whatever the task required of them — choosing among written endings of varying length or estimating the length of the ending in real time — subjects are surprisingly good not only at differentiating the sentences that are ending from those that are continuing but also at differentiating those that continue for three more words from those that continue for six more words.

A comparison of the estimated durations proposed by the subjects at the 100% gate with the actual physical durations of the endings at that gate shows a systematic overshooting on the part of the subjects. Thus, for the zero-word endings, the mean estimated time was 379 msecs instead of 0 msecs; for the three-word endings, the estimated time was 1218 msecs instead of 688 msecs; and for the six-word endings the estimated time was 1784 msecs instead of 1068 msecs. This overshooting could be due to the difficulty of the task (subjects usually found it hard to do) or to some aspect of the response mechanism (some constant delay in pressing the key, among other things). What is interesting, however, is that subjects separate the zero-word endings from the six-word endings by a mean difference of 1405 msecs (the actual difference was 1068 msecs). In other words, there is enough prosodic information in the sentence fragment to indicate that in some cases the end of the sentence is immediate and that in other cases it is situated some one and a half second later — a very long time indeed when one is waiting to press a key!

An acoustic analysis of the test sentences

We undertook an acoustic analysis of the 32 test exemplars in order to determine whether three acoustic variables — fundamental frequency, amplitude, and duration — were correlated with the results obtained in the two experiments. We hoped that some relationship would be found but realized that this was not a necessity. As was stated earlier, the ability to predict the length of a sentence ending might involve the complex integration of a number of acoustic variables (as in the perception of stress, for example), and correlating our data with simple acoustic measures might not reveal this integration. As we will see below, however, the relationship between the data and the acoustic variables appears to be quite straightforward.

We used Maeda's ANALYS program (Maeda 1976) to obtain the amplitude envelope and the Fo contour of each of the 32 sentence exemplars as well as individual Fo, amplitude, and time values at chosen points during each exemplar. In Figure 5 we present the envelopes and contours of two exemplars: the zero-word-ending version of sentence 1 (top part) and the six-word-ending version of the same sentence (bottom part). A number of measures can be calculated from such displays (such

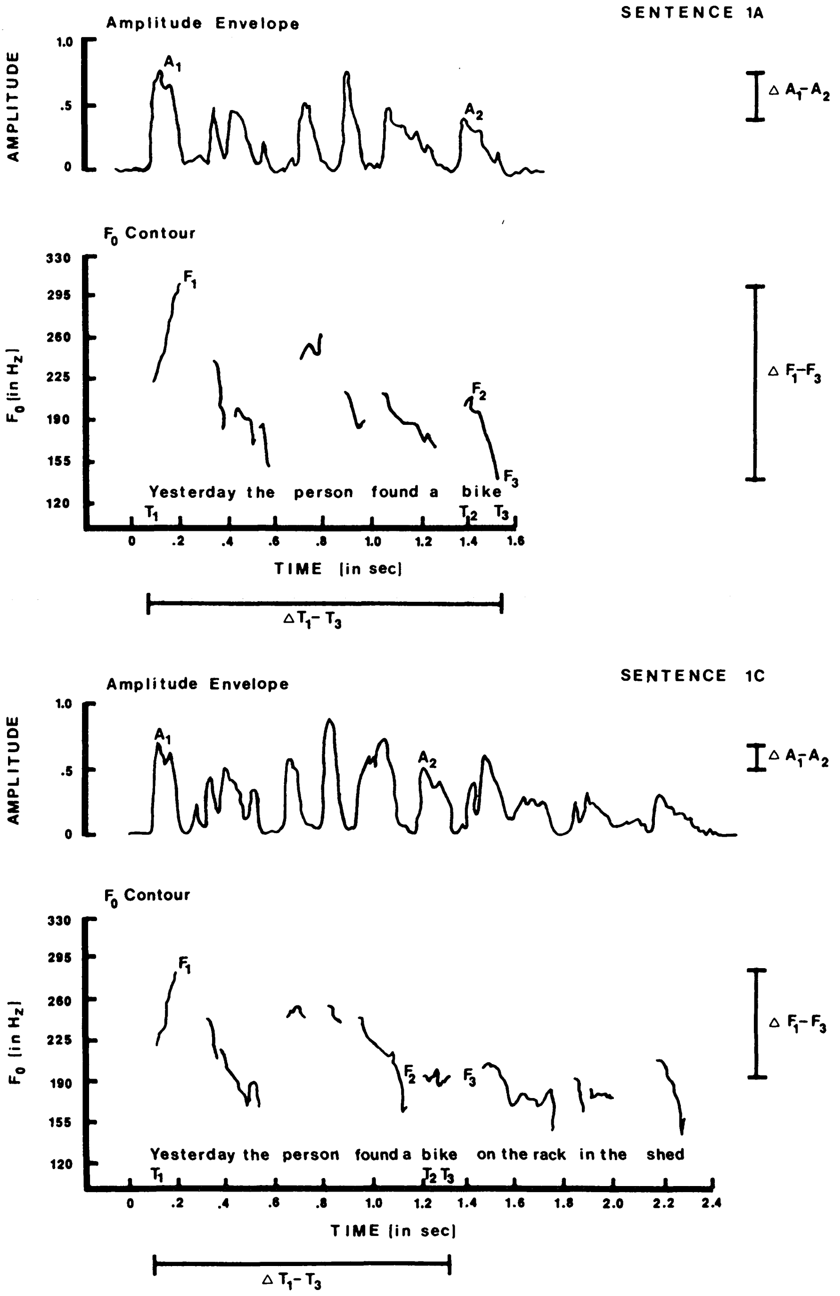

Figure 5. *The amplitude envelope and Fo contour of two sentence exemplars (0-word ending and 6-word ending) as produced by the ANALYS program (Maeda 1976: n. 2). Each vertical and horizontal bar represents the difference between an acoustic value at the beginning of the sentence and its counterpart at the end of the object noun*

as the declination and the amplitude slopes), but we decided to use (with one exception) the DIFFERENCE between the value of that same variable at another point in the sentence.

Two domains of measurement were defined: the sentence-fragment domain which included all the words of the sentence up to and including the stimulus word (but not the various endings), and the stimulus-word domain. For each domain and each sentence, we computed three acoustic values: an Fo value, an amplitude value, and a duration value. In the sentence-fragment domain, the Fo value was the difference between the Fo of the peak of the first word (F_1 in Figure 5) and the Fo at the end of the stimulus word (F_3). This difference is illustrated by vertical bars labeled $\triangle$ F_1–F_3 in Figure 5. The amplitude value was the difference between the amplitude peak of the first word (A_1) and the amplitude peak of the stimulus word (A_2). This difference is also illustrated by vertical bars labeled $\triangle$ A_1–A_2 in Figure 5. And the duration value was the difference between the onset time of the first word of the sentence (T_1) and the offset time of the stimulus word (T_3). This difference is illustrated by the horizontal bars labeled $\triangle$ T_1–T_3 in Figure 5. In the stimulus-word domain, the Fo value was the difference between the first peak (or first dip) of the word (F_2) and the lowest (or highest) point reached by the Fo at the end of the word (F_3).[4] The amplitude value was the relative amplitude of the peak of the stimulus word (A_2) and therefore was not a difference; and finally, the duration value was the actual duration of the word (T_2–T_3). The mean values obtained in this way for the three acoustic variables in the two domains of measurement are presented in Table 1.

Table 1. *Mean values of three acoustic variables in two domains of measurement. Except for the mean amplitude of the stimulus word, all other values are differences between a value at the beginning of the domain and one at the end. Each mean is based on eight observations*

Domain of measurement	Variable	Ending type 0-word	3-word	6-word	9-word	$F(3, 21)=$
Sentence fragment	amplitude (relative)	0.43	0.26	0.22	0.24	8.74, $p<0.01$
	Fo (Hz)	148	98	76	81	8.59, $p<0.01$
	duration (Msec)	1548	1353	1313	1301	47.15, $p<0.01$
Stimulus word	amplitude (relative)	0.41	0.48	0.54	0.50	8.43, $p<0.01$
	Fo (Hz)	77	20	8	8	15.46, $p<0.01$
	duration (Msec)	320	190	181	181	130.82, $p<0.01$

Before correlating the sets of values with the experimental data, we undertook a preliminary study of the acoustic values. We ran one-way analyses of variance on the sets of values for the two different measurement domains. Thus, for example, we took the Fo values in the sentence-fragment domain and tested for a difference between the eight values obtained for each of the four types of sentences. The six ANOVAs that were run in this way all produced significant main effects for *ending* (see Table 1). For example, in the sentence-fragment domain, we found that as the ending was lengthened, the difference in Fo values (F_1–F_3) decreased significantly. This is clearly illustrated in Figure 5, where we see that the F_1–F_3 difference is much larger in Sentence 1A (a zero-word-ending sentence) than in Sentence 1C (a six-word-ending sentence). A main effect was also found for amplitude, indicating that as the ending was lengthened, the difference in amplitude also decreased significantly (see the A_1–A_2 differences in Figure 5). And in the same domain, a main effect was found for duration, indicating that as the ending was lengthened, the duration of the sentence fragment also decreased significantly. As can be seen in Table 1, these three main effects were also found in the stimulus-word domain. (Note that the amplitude values increase in this domain, as they are not based on a difference but on actual relative amplitudes.)

A posteriori tests on all six analyses of variance revealed in each case a significant difference between the values of the zero-word endings and those of the three-, six-, and nine-word endings. However, only a trend was found in the difference between the values of the three- and six-word endings. As for the values of the nine-word endings, we can see in Table 1 that they are practically identical to those of the six-word endings, and in three cases they are actually situated between the values of the three- and six-word endings. Thus, for example, in the sentence-fragment domain, the mean Fo value for the nine-word ending is 81 Hz, whereas that for the six-word domain is 76 Hz and that for the three-word ending is 98 Hz. This similarity in the acoustic values of the sentences containing six- and nine-word endings explains in large part the similar estimates given by subjects in experiments 1 and 2 (see Figures 1 and 4).

We also tested whether the Fo value of the first word of the sentence (F_1) and the value of the very last word of the sentence were influenced by the total length of the sentence. We found that this was not the case. Mean Fo values on the first word in the zero-, three-, six-, and nine-word-ending sentences were similar: 299, 288, 286, and 289 Hz respectively [$F(3, 21) = 2.31$, N.S.], as were the mean values at the end of the last word of the sentence: 151, 148, 156, and 153 Hz respectively [$F(3, 21) = 0.26$, N.S.]. This means that if subjects rely on the Fo to make their estimates of sentence length, they cannot use the height of the first peak, as in this case it is not a good indicator of sentence length.

In Table 2 we present the Pearson product–moment correlations between the data obtained at the 100% gate in experiments 1 and 2 and the values of three acoustic variables obtained from each of the two domains of measurement. Each correlation is based on 32 observations. We have also included the multiple regression coefficients obtained when all three predictor variables are combined to predict the data. A number of points emerge from an examination of the table. First, all but one of the correlations are significant, indicating that the acoustic variables do indeed appear to be reliable predictors of the data. We can hypothesize from this that when making their judgments of sentence length, listeners probably do base themselves on something like the difference in the acoustic values between one point in the sentence and another. Second, the Fo and the duration values are better predictors of the data than are the amplitude values. Similar findings were found by Streeter (1978) when accounting for the prosodic disambiguation of sentences. Third, the multiple regression coefficients are all very high (they range from 0.64 to 0.83), indicating that much of the variance in the data can be accounted for by combining the three acoustic measures. Fourth, the values obtained from the word domain are usually better predictors of the data than are the values obtained from the sentence-fragment domain. This can be explained in one of two ways. Either the subjects wait for the potentially last word to make their estimates (as was suggested earlier) and base themselves only on the prosodic information of the word, or the subjects take into account the acoustic information of both the sentence

Table 2. *The Pearson product–moment correlations between the data obtained in experiments 1 and 2 and three acoustic variables in two different domains of measurement. Each correlation is based on 32 observations. * indicates a 0.05 level of significance and ** a 0.01 level. Multiple regression coefficients based on the three acoustic variables are given to the right of the table*

Dependent variables	Domains of measurement	Predictor variables Fo	Amplitude	Duration	Multiple R
Estimated number of words (exp. 1)	sentence fragment	−0.60**	−0.49**	−0.65**	0.71
	stimulus word	−0.75**	0.42*	−0.81**	0.83
Estimated time (exp. 2)	sentence fragment	−0.55**	−0.48**	−0.56**	0.64
	stimulus word	−0.67**	0.29	−0.61**	0.69

fragment and the potentially last word but put a bit more weight on the acoustic values of the word. The former explanation, where only the word domain counts, fits better with the fact that subjects cannot seem to differentiate the various sentence endings at the beginning of the stimulus word. A direct test of this explanation would be to excise the stimulus words from the speech stream and to ask subjects to predict the length of the remaining part of the sentence. If they manage to do this, then the acoustic information 'window' is rather short and is confined to the stimulus word. If the subjects can no longer do the task, however, then the apparent stimulus-word superiority effect reflects only that subjects take into account the acoustic information of both the sentence fragment and the word but put a bit more weight on the latter. A fifth and final point is that the acoustic measures predict the data of experiment 1 slightly better than that of experiment 2, whether in the sentence-fragment domain or the word domain. This probably only reflects the extra noise produced by the inherent difficulty of the key-press task. The multiple R values, although lower, are still quite high (0.64 and 0.69). In sum, therefore, simple acoustic measures of the sentences are good predictors of the experimental data, suggesting that subjects may well be using some basic analysis of change in Fo, duration, and amplitude to make their predictions of sentence length.

General discussion

The present study has confirmed that listeners are very good at using prosodic information to determine whether a sentence is over or not when listening to the potentially last word of a sentence. More surprisingly, however, the study has shown that listeners are surprisingly proficient at estimating how much longer a sentence will continue when it is not over, and this by relying solely on prosodic cues. This predictive ability is important because, if used during on-line processing, it can indicate the length of the unit remaining and can therefore help delineate the end of the sentence and hence the beginning of the next one. This in turn will facilitate parsing and end-of-sentence packaging as well as helping the listener prepare a response if he or she has decided to intervene.

An interesting finding in the study has been that prediction is quite random at the beginning of the potentially last word, whereas it is very good at the end. A control study showed that this was not due to a 'wait and see' strategy on the part of the subjects or to the repetitive nature of the paradigm. Two explanations were retained to account for this phenomenon. The first is that the information needed for prediction of this kind is only carried by the stressed syllables and so subjects have to wait for a large part of the last noun to be given to them to make their estimates. The second is that predicting the length of the remaining part of the sentence can only take place on the potentially last word, and hence subjects have to wait for that word to make their prediction. Only further experimentation will show which of these two explanations is correct.

An encouraging finding in the study has been that the results are not response-specific. Subjects produced very similar patterns of results with two very different response tasks, a multiple-choice task and a key-press task. In each case, they could not differentiate the various ending types at the beginning of the potentially last word but started doing so as they progressed through the word and were surprisingly accurate on three of the four endings at the end of the word.

Finally, it was found that the results were well predicted by three acoustic measures: Fo, duration, and amplitude. Especially interesting was the fact that the acoustic measures of the potentially last word itself were good predictors of the data, indicating the possibility that the only information necessary for prediction is carried by that word.

Of course, this study asks many more questions than it answers. For example, can a subject predict the length of the sentence at any point within the sentence or must he or she wait for a stressed syllable? If a stressed syllable is needed, how many stressed syllables are needed before accurate prediction can take place? Is there a cumulative effect of stressed syllables or is each stressed syllable treated independently of the preceding stressed syllables in the prediction process? Finally, must the subject reach a critical word (the potentially last word of the sentence) to have access to the relevant prosodic information and/or be able to predict the length of the remaining part of the sentence? These and other questions will have to await other studies.

An interesting proposal that has been made (J. Gee, personal communication) is that the prediction system is basically binary and contains two features. The first is [±continuation] and the second is [± long ending]. Basically, the listener first decides whether the sentence is continuing or not. If he or she decides that it is continuing, then a judgment is made on the length of the ending. Thus, two binary features lead to a three-way contrast, similar in many ways to that found in phonology with the [± high vowel], [± low vowel] features. This proposal could explain the fact that subjects differentiated three of the four endings: the zero-word ending $\begin{bmatrix}-\text{continuation}\\-\text{long ending}\end{bmatrix}$, the three-word ending $\begin{bmatrix}+\text{continuation}\\-\text{long ending}\end{bmatrix}$, and the six-word ending $\begin{bmatrix}+\text{continuation}\\+\text{long ending}\end{bmatrix}$, but not the nine-word ending, which is also $\begin{bmatrix}+\text{continuation}\\+\text{long ending}\end{bmatrix}$. However, the present study is not a good test of this model, as the acoustic measures showed the last two endings (six and nine words) to be similar in nature. To test it, one will need to vary ending length from zero to six words, in steps of one word, for instance. If subjects can differentiate the endings, then a two-feature prediction system will not be adequate; if, however, the results fall into three main groups, then the proposal may well be correct.

This last topic raises another interesting question: what is the unit of length that subjects are using in their prediction? Is it the syllable, the word, the phrase? Or is it the stressed syllable or the phonological phrase? In the present study, we adopted the word as the unit of prediction, but an increase in the number of words was also accompanied by an increase in the number of syllables and phonological phrases. A future study will have to decouple these units in order to determine which unit is used by subjects when predicting the length of the remaining part of the sentence.

Of course, all of the above questions will have to be asked not only about read-out speech but also about spontaneous speech. In the present study, our sentences were constructed and recorded in a laboratory environment, and hence we may ask if similar results would have been obtained with sentences that were produced spontaneously. To answer this question we would need to find sentences in spontaneous interviews that are similar to the ones we used here (that is, sentences with adverbial or prepositional phrases) and to splice them at the same critical break. If the auditory effect produced by poorly spliced radio interviews is any evidence, we should obtain results that are similar to those of the present study. It is interesting to note that Miller et al. (1982) have found a significant mean correlation of 0.33 between the length of a run in spontaneous speech (that is, the stretch of speech between two pauses) and the articulation rate of that run. This finding is important, as duration (and hence rate) in our study was highly related to estimated sentence length.

In addition to answering some questions but also raising many other questions about the prediction of sentence length, this study should be seen as yet another invitation to examine the time course of prosodic information in the speech stream. The gating paradigm, among other tasks, can be used to determine how much prosodic information and which type of information is available at a particular point in time in the utterance, and how this information can be used to help predict what is coming up. Examples of questions that can be answered with this approach are, how early in a tag question (e.g. 'This isn't a new car, is it?') does the listener know that he or she is dealing with such a question? how about in a yes–no question (e.g. 'This is a new car?') ? does the listener know that a yes–no question is being asked only on the last stressed word, or does some information exist before that point? what about sentences with emphatic stress (e.g. 'This is a *blue* car?') Besides these specific questions, the whole time course of the prosody that is used to convey important semantic information, such as anger, emotion, doubt, surprise, can be studied in this way, as can the time course of prosody in the sign languages of the deaf — an area of research that is practically unexplored.

This use of the gating paradigm cannot show that listeners actually do use prosodic information to predict upcoming events during the on-line processing of the utterance. Other paradigms will need to do that. But it can help specify the amount of prosodic information that is available to the listener at a particular point in time and whether this information can be used by the listener in the prediction of upcoming events. This in itself is an important step in helping us understand the on-line processing of language and the role that prediction plays during this activity.

Received 26 January 1983
Revised version received
8 December 1983

Department of Psychology
Northeastern University
360 Huntington Avenue
Boston, Mass. 02115
USA

Appendix

The sentence exemplars used in experiments 1 and 2. Zero-word-ending exemplars are marked by a slash immediately after the stimulus word (underlined); three-, six-, and nine-word-ending exemplars are also marked by slashes at corresponding distances from the stimulus word.

Earlier my sister took a dip/ in the pool/ at the club/ on the hill/
Earlier the person chased the cat/ from the couch/ in the room/ of the house/
Yesterday my brother found a dog/ on the bench/ in the park/ of the town/
Yesterday the toddler threw a cup/ at the child/ in the yard/ of the school/
Recently my father climbed a peak/ of the range/ in the north/ of the Alps/
Recently the trainer bought a colt/ from the man/ from the ranch/ on the hill/
Yesterday my sister made a cake/ for the fair/ at the school/ on the hill/
Yesterday the person found a bike/ on the rack/ in the shed/ of the yard/

Notes

* This study was supported in part by grants from the Department of Health and Human Services (RR 07143 and NS 14923). The author would like to thank Arthur Wingfield for his enthusiastic comments at the beginning of the study, Jim Gee, Leah Larkey,

Joanne Miller, Carlos Soares, and an anonymous reviewer for their many insightful comments on the draft of the study, and Peter Eimas and Steve Harkins for their help with the statistical analysis. Special thanks go to Dr. Kenneth Stevens for making available the facilities of the Speech Communication Laboratory at MIT, Dr. S. Maeda for supplying the ANALYS program, and S. Hawkins, J. Shepard-Kegl, and J. Gee for their stress and intonational analysis of the test sentences. Requests for reprints should be sent to François Grosjean, Department of Psychology, Northeastern University, Boston, MA 02115.

1. An anonymous reviewer has suggested that there may be a confound between the type of sentence used and the size of the presentation set. That is, sentences with a zero-word ending were represented by a larger presentation set (eight to ten gates depending on the word in question, from which five gates were analyzed), whereas sentences with a three-, six-, or nine-word ending were represented by a smaller presentation set (five to six gates, depending on the word). This, according to the reviewer, resulted in a higher accuracy in responses and higher confidence ratings for the zero-word endings. The following points can be made to allay the reviewer's fears: (a) a close examination of the results given at the last gate of each type of sentence (see Figure 1 for example) does not reveal that Ss are any more accurate with zero-word endings than with three- or six-word endings. In fact, the reverse seems to be true; (b) a control study described in the next pages shows that Ss presented with the last gate only (and therefore not able to work their way through to the last gate) give identical results to those obtained in the experiment; (c) along the same line, a study by Cotton and Grosjean (1984) in another domain — word recognition — shows that Ss given a single gate taken from a presentation set give identical results to those Ss who work their way through a presentation set. From this we conclude that for the latter Ss the answer given to a preceding gate has little — if any — impact on the answer they will give to the gate that is being presented to them.
2. A simple one-way ANOVA on the results obtained solely from the successive presentation control study confirm this. No main effect was found for *ending* [$F(3, 21) = 2.67$, N.S.] and a *post hoc* Tukey HSD shows that none of the means are significantly different from one another.
3. We should note that at no point during the sentences, prior to the stimulus word, could Ss have felt they had reached the potentially last word of the sentence. An examination of the following examples illustrates this very well:

   Yesterday my sister made a *cake*
   Earlier my sister took a *dip*
   Yesterday the person found a *bike*

4. Care was taken to retain the sign of the difference, indicating thereby a rising, steady, or falling contour.

References

Bever, T. (1970). The cognitive basis for linguistic structures. In J. Hayes (ed.), *Cognition and the Development of Language*. New York: Wiley.

Brazil, D. (1981). The place of intonation in a discourse model. In M. Coulth and M. Montgomery (eds.), *Studies in Discourse Analysis*. London and Boston: Routledge and Kegan Paul.

Breckenridge, J. (1978). Declination as a phonological process. Unpublished manuscript, Department of Linguistics and Philosophy, MIT.

Clark, H., and Clark, E., (1977). *Psychology and Language*. New York: Harcourt Brace Jovanovitch.

Comrie, B. (1981). *Language Universals and Linguistic Typology*. Chicago: University of Chicago Press.

Cotton, S., and Grosjean, F. (1984). The gating paradigm: a comparison of successive and individual presentation formats. *Perception and Psychophysics* 35, 41–48.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception and Psychophysics* 20, 55–60.

—, and Darwin, C. (1981). Phoneme-monitoring reaction time and preceding prosody: effects of stop closure duration and of fundamental frequency. *Perception and Psychophysics* 29, 217–224.

—, and Fodor, J. (1979). Semantic focus and sentence comprehension. *Cognition* 7, 49–59.

—, and Foss, D. (1977). On the role of sentence stress in sentence processing. *Language and Speech* 20, 1–10.

Fodor, J., Bever, T., and Garrett, M. (1974). *The Psychology of Language*. New York: McGraw Hill.

Frazier, L., and Fodor, J. D. (1978). The sausage machine: a new two-stage parsing model. *Cognition* 6, 291–335.

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

Kimball, J. (1973). Seven principles of surface structure parsing in natural language. *Cognition* 12, 15–47.

Kirk, R. (1967). *Experimental Design: Procedures for the Behavioral Sciences*. Belmont, Calif.: Brooks/Cole.

Lehiste, I. (1973). Phonetic disambiguation of syntactic ambiguity. *Glossa* 7, 107–121.

Maeda, S. (1976). A characterization of American English intonation. Unpublished Ph.D. dissertation, MIT.

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

—, and Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology* 10, 29–63.

Martin, J. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behaviour. *Psychological Review* 79, 487–509.

Meltzer, R., Martin, J., Mills, C., Imhoff, D., and Zohar, D. (1976). Reaction time to temporally-displaced phoneme targets in continuous speech. *Journal of Experimental Psychology: Human Perception and Performance* 2, 277–290.

Miller, J., Grosjean, F., Gee, J., and Lomanto, C. (1982). Changes of speaking rate in spontaneous speech. Unpublished report, Northeastern University.

Ohman, S. (1966). Perception of segments of VCCV utterances. *Journal of the Acoustical Society of America* 40, 979–988.

Pollack, I., and Pickett, J. (1963). The intelligibility of excerpts from conversation. *Language and Speech* 6, 165–171.

Scott, D. (1982). Duration as a cue to the perception of a phrase boundary. *Journal of the Acoustical Society of America* 71, 996–1007.

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.

Streeter, L. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America* 64, 1582–1592.