Published in final edited form as:

J Mem Lang. 1985 April ; 24(2): 210–231. doi:10.1016/0749-596X(85)90025-7.

Interaction of Knowledge Sources in Spoken Word Identification

Aita Salasoo and David B. Pisoni

Indiana University

Abstract

A gating technique was used in two studies of spoken word identification that investigated the relationship between the available acoustic–phonetic information in the speech signal and the context provided by meaningful and semantically anomalous sentences. The duration of intact spoken segments of target words and the location of these segments at the beginnings or endings of words in sentences were varied. The amount of signal duration required for word identification and the distribution of incorrect word responses were examined. Subjects were able to identify words in spoken sentences with only word-initial or only word-final acoustic–phonetic information. In meaningful sentences, less word-initial information was required to identify words than word-final information. Error analyses indicated that both acoustic–phonetic information and syntactic contextual knowledge interacted to generate the set of hypothesized word candidates used in identification. The results provide evidence that word identification is qualitatively different in meaningful sentences than in anomalous sentences or when words are presented in isolation: That is, word identification in sentences is an interactive process that makes use of several knowledge sources. In the presence of normal sentence context, the acoustic–phonetic information in the beginnings of words is particularly effective in facilitating rapid identification of words.

It is now well documented that both acoustic–phonetic information from the speech signal and other nonsensory sources of knowledge contribute to spoken word identification (see Bagley, 1900; Cole & Rudnicky, 1983). The listener's knowledge of morphology, syntax, and semantics may be collectively labeled *context*. The problem of how context is used to support perception and comprehension is one of the most important questions in the field of language processing today (e.g., Cole & Rudnicky, 1983; Grosjean, 1980; Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980; Stanovich & West, 1981; Swinney, 1979).

Early studies of context effects in speech processing employed degraded stimuli (e.g., Miller, Heise, & Lichten, 1951; Miller & Isard, 1963). With impoverished speech signals such as sentences presented against high levels of noise, listeners can extract the linguistic content of the message if they have access to normal semantic and syntactic information. When these top-down knowledge sources are experimentally removed or modified in some way (Miller & Isard, 1963), listeners' identification suffers substantially. If we consider the contribution of distortion and masking of the speech signal by miscellaneous sources of noise, for example, coughs, traffic, faulty telephone lines, then it follows that non-acoustic knowledge sources also contribute to the identification of words in normal fluent speech processing situations.

More recently, a variety of tasks have been used to address the mechanisms of context effects. One particular approach has been to observe the effect of various context

manipulations on measures that index the listener's relative dependence on bottom-up acoustic–phonetic information.

In this spirit, Marslen-Wilson and Tyler (1980) found that words in normal sentences were recognized before half of their acoustic–phonetic signal had become available to listeners. They also observed that identification times (to a specified word target) increased in anomalous sentences that had no coherent semantic interpretations, compared to normal sentences. Identification times were longest for words in scrambled lists of unrelated words. Several studies have used a gating paradigm, which measures the minimum acoustic–phonetic input required for word identification (Grosjean, 1980; Cotton & Grosjean, 1984). Listeners in these studies required less stimulus information to identify words in sentence contexts than when the same words occurred in isolation. The locus and mechanisms producing these context effects are not well understood, although they are the focus of much current research (cf. Cole & Rudnicky, 1983).

In this paper, we investigate the sources of knowledge that are employed in spoken word identification. We will assume that at some stage in perceiving and processing the speech input, words are identified or recognized by listeners. The word identification process involves a number of component stages including word recognition, lexical access and retrieval, and response execution. We will adopt the term *word identification* to stand for the conscious belief (and a response contingent on that belief) that a particular word has just occurred. We will reserve the term *word recognition* for the results of the low-level sensory pattern-matching process that is assumed to occur upon hearing a spoken word. By *lexical access* we mean contact of an internal representation derived directly from the speech input with a lexical representation (i.e., a word) in memory and retrieval or activation of that item in working memory (Pisoni, 1981).

In recent years, questions about the mechanisms responsible for context effects in speech processing have focused on the issue of autonomous versus interactive processing (Cairns, 1982; Marslen-Wilson & Tyler, 1980; Norris, 1982a, 1982b; Swinney, 1982; Tyler & Marslen-Wilson, 1982a, 1982b) and on the special status given to word-initial phonetic segments in lexical access (Cole & Jakimik, 1980; Garrett, 1978; Cairns, 1982). According to the autonomy principle (e.g., Swinney, 1982, p. 164), lexical processing consists of “a set of isolable, autonomous substages, where these substages constitute domain specific processing modules.” Of immediate interest to us is the a priori assumption made by several investigators that lexical access is autonomous or context independent (e.g., Forster, 1979; Marslen-Wilson & Tyler, 1980; Swinney, 1982). In autonomous accounts of spoken word identification, the effects of higher-order context are assumed to be postlexical in nature (Forster, 1979).

In the original account of cohort theory, Marslen-Wilson and Welsh (1978) assumed that the initial set (or cohort) of word candidates is fully determined by the word-initial acoustic–phonetic input alone (see also Tyler & Marslen-Wilson 1982a, 1982b). According to Marslen-Wilson's “principle of bottom-up priority” (e.g., Marslen-Wilson & Tyler, 1980; Tyler & Marslen-Wilson, 1982b) the cohort-establishing processes (involved in lexical access as we have defined it) are viewed as context independent, that is, autonomous, using only word-initial acoustic–phonetic information. Once the set of word-initial cohorts is activated, subsequent bottom-up acoustic–phonetic information and all other sources of information (including syntactic and semantic constraints) are used to deactivate incompatible word candidates. According to this view, a word is identified “optimally” at the point where it becomes “uniquely distinguishable from all of the other words in the language beginning with the same sound sequence” (Tyler & Marslen-Wilson, 1982b, p. 175).

The autonomy principle gives special status to the initial acoustic–phonetic information in the speech signal for directing the word identification process. The cohort theory makes strong claims about the role of word-*initial* acoustic–phonetic information, namely, that an entire set of lexical candidates that share word-initial phonetic information with the spoken stimulus word is directly activated during word identification. Similarly, in Forster's (1976) autonomous search model, a master file in the lexicon contains all the phonetic, syntactic, and semantic information associated with a word token that is subjected to postaccess comparison and identification decision processes. Entry to the master file proceeds, in the case of spoken language input, only via a peripheral phonetic file based on acoustic–phonetic information in the signal. As Swinney (1982) has pointed out, in both accounts, it is assumed a priori that initial processing is autonomous.

Tyler and Marslen-Wilson (1982b) have located the effects of context in postlexical decision stages of processing, almost by default. This conclusion is based on observations that word-initial fragments are influential in postperceptual tasks. For example, in both visual and auditory domains, production and memory for words when their initial segments are presented are superior to performance based on their final segments (e.g., Bruner & O'Dowd, 1958; Nooteboom, 1981). However, the possibility of perceptual context effects in continuous spoken word identification cannot be ruled out by these data.

Grosjean's (1980) recent analysis of listeners' responses in an identification task constituted the first attempt to operationalize Marslen-Wilson's concept of a word-initial cohort. Grosjean used a gating paradigm in which signal duration was varied: Listeners tried to identify a target word after fragments of the speech signal had been presented. Incorrect responses included not only acoustically similar words but also word candidates guided by semantic constraints and word frequency. Grosjean (1980) interpreted his response data as evidence against the claim that only acoustic–phonetic information controlled the distribution of lexical candidates generated before a word is identified. Instead, Grosjean (1980) suggested that *both* acoustic and nonacoustic sources of knowledge interact to select potential word candidates in lexical access. This view is similar to the interactive logogen model proposed by Morton (1969, 1979).

Using a task based on Grosjean's gating technique, we examined the interaction of contextual and sensory information in the identification of content words in sentences. We replaced multiple word targets in sentences with envelope-shaped noise and asked listeners to identify all the words. Our gating task is an "off-line" procedure that employs unusual sentence stimuli. Nevertheless, the task appears to be sensitive to the amount of sensory input available at various points during the time course of spoken word identification. It is possible that additional strategies not used normally in speech comprehension, such as hypothesis testing and guessing, may also be reflected in our task. We believe, however, that our results are compatible with those from "on-line" techniques and therefore can provide converging evidence about the availability of various knowledge sources used to identify words in spoken sentences.

In two gating experiments reported below, we investigate the interactive assumption that normal spoken word identification processes require the presence of semantic and syntactic context. More specifically, we focus on the a priori nature of the autonomy assumption in word identification and the special status given to word-initial acoustic–phonetic information in cohort theory.

Experiment 1

The present study used a gating paradigm to investigate the knowledge sources employed in the identification of words in spoken sentences. Three questions were addressed: First, is

word-initial acoustic–phonetic information obligatory for word identification in sentences? Second, how does the reliance on acoustic–phonetic information change across normal, meaningful sentences, and syntactically normal, but semantically anomalous sentences? And, finally, how is the distribution of incorrect responses to “word-initial” or “word-final” information related to the amount of signal duration required for identifying spoken words?

Our terminology and procedure differ somewhat from previous studies using various gating techniques.¹ In our task, subjects were required to identify content words in short sentences after each presentation of the sentences. On the first trial, the waveform of each target word was replaced completely by envelope-shaped noise. This noise effectively removed all segmental acoustic–phonetic cues, while at the same time preserving the natural prosodic and duration information. On each consecutive trial, 50-millisecond increments of the original waveform replaced selected parts of the noise in each target word. The number of consecutive 50-millisecond increments increased on successive repetitions of the sentence until, on the final trial, the entire waveform of the original word was presented. Two aspects of our procedure are novel. First, in constructing the stimuli, we replaced the non-presented part of the signal with envelope-shaped noise instead of with silence. Second, we used multiple target words in sentences to simulate the demands of normal, continuous word identification in processing fluent speech. This procedure contrasts with the use of a single target item that has often been the final word in a sentence (e.g., Grosjean, 1980; Cotton & Grosjean, 1984). Our use of the term “gate” refers to the duration of the presented segment of the intact speech signal for each of the target words.

Method

Subjects—The subjects were 194 introductory psychology students who received course credit for their participation. All subjects were native speakers of English with no reported hearing loss. None of the subjects had participated in previous experiments using speech or speech-like materials.

Materials—From two sets of experimental materials, three context conditions were generated: words in meaningful sentences, words in anomalous sentences, and the same words presented in isolation. Eight Harvard psychoacoustic sentences (Egan, 1948) were chosen for the meaningful context condition. These sentences, for example, “The stray cat gave birth to kittens,” cover a range of syntactic structures and are balanced according to word frequency and phonological density counts in English usage. The Harvard sentences are typical of active declarative sentences in English; they are meaningful and syntactically normal. The second context condition consisted of eight sentences selected from a set of materials originally developed for use in the evaluation of the intelligibility of synthesized speech (Nye & Gaitenby, 1974; Pisoni, 1982). These materials, known as the Haskins sentences, are syntactically normal and contain high-frequency words. Unlike the Harvard sentences, however, they are semantically anomalous, for example, “The end home held the press.” As such, the Haskins sentences represent a class of semantically impoverished sentences. The two sets of materials will be referred to as meaningful (Harvard) and anomalous (Haskins) sentences, respectively.

¹Pollack and Pickett (1963) presented listeners with single presentations of words that had been excised out of spoken sentences; they called these “gated stimuli.” Correct identification of such stimuli was often impossible. Other researchers have used truncated nonsense syllables as a stimuli (Ohman, 1966) or isolated gated word stimuli (Grosjean, 1980; Cotton & Grosjean, 1984).

All the content words from the meaningful and anomalous sentences served as targets.² In the third condition, the target words were excised from the intact spoken sentences and presented in isolation. This condition serves as a control for the contribution of any sentence context per se in the word identification process (Miller et al., 1951).

Two properties of the target words were varied orthogonally: first, the amount of acoustic–phonetic information in the waveform, defined by gate duration; and second, the location of that information within the word, defined by gating direction. Stimulus duration was varied in 50-millisecond increments between successive trials. The two levels of gating direction were forward, with increasing amounts of signal, left-to-right, from the word beginning, and backward, with increasing amounts of signal, right-to-left, from the end of the word.

Audio tapes of the original sentences, read by a male speaker, were sampled at 10 kHz, low-pass filtered at 4.8 kHz, digitized, and stored on disk by a PDP-11/34 computer. The beginnings and endings of target words were located with a digital waveform editor. The gated conditions of the target words in each sentence were produced by simply replacing the appropriate number of consecutive 50-millisecond intervals with envelope-shaped noise (Hori, House, & Hughes, 1971). For each digital sample of the waveform, the direction of the amplitude was reversed at random while the absolute value of the amplitude and the RMS energy were preserved. This procedure maintained the amplitude and durational cues of the original speech waveform, while at the same time obliterating the fine spectral information (i.e., formant structure) needed for segmental identification.

For each original sentence, two sequences of experimental sentences were produced, one each for the forward and backward gating conditions. In both sets of materials, the first and last trials were identical: On the first trial, all target words were completely replaced by noise, while on the last trial the original intact spoken sentence was presented. The remaining trials contained acoustic–phonetic information increasing in 50-millisecond increments from the beginning (for a forward-gated sentence) or ending (for a backward-gated sentence) of each target word. Figure 1 shows spectrograms of the first, third, fifth, and last trials (from top to bottom) of forward-gated and backward-gated sequences of one of the meaningful sentences used in the experiment. The isolated word conditions were created by excising words from the gated sentences (cf. Pollack & Pickett, 1963). Forward-gated and backward-gated sequences were created separately for each target word in isolation. Each isolated gated word presentation was treated as a trial.

Sixteen experimental tapes were created from the digitally stored stimuli using a 12-bit D/A converter and a Crown 800 Series tape recorder. For each gating direction two counterbalanced random orders of the eight sentence (or word) sequences were generated in each context condition. Each sequence progressed from the stimulus with shortest gate duration to the stimulus with the longest gate duration.

Procedure—Small groups of subjects were tested simultaneously in a sound-treated experimental room. Each group heard one experimental tape at 77 dB SPL peak levels over TDH-39 matched and calibrated earphones. Between 20 and 26 subjects heard each context type for each gating direction.

Subjects were told that they would hear a single sentence or word on each trial, and that each stimulus would be repeated, becoming clearer on each subsequent presentation.

²Behavioral evidence suggests that function words may be identified with different processes and knowledge sources than content words (Garrett, 1978; Marslen-Wilson & Tyler, 1980; Salasoo & Pisoni, 1981). Therefore, function words in the experimental sentences were not treated as target items, but, instead, remained acoustically intact across all conditions.

Subjects were instructed to write down the word or words they heard after each presentation. Subjects were encouraged to guess if they were not certain of a particular word or words. The experimenter stopped the tape recorder manually via a remote control unit after each trial and continued only when all subjects had finished writing their responses on prepared answer sheets. In the isolated word condition, the tape ran without interruption; the intertrial interval was 3 seconds long and cue tones were used to signal the start of a new stimulus sequence.

In the sentence conditions, response sheets contained the function words in their correct locations along with separate blank spaces for the content words in each experimental sentence. Subjects, therefore, had some access to top-down knowledge provided by the information on the answer sheets, for example, the possible form class of words following function words and the number of words in the sentences. Subjects were required to respond to each word after each stimulus presentation with either a word or an “X” if they could not identify a word.

Results and Discussion

Two types of dependent measures were obtained. First, we computed the “identification point” for all target words. This was defined as the duration of the signal present on the trial during which the word was first correctly identified and then continued to be identified correctly thereafter by a subject.³ Second, we analyzed subjects’ incorrect word responses in order to examine the response distribution of word candidates that was generated in a given gating condition before a given target word was identified correctly. We considered these response distributions to be an empirical measure of the structure of the underlying cohort (cf. Grosjean, 1980). That is, we assumed that the incorrect word responses before a word was identified would reflect the pool of potential word candidates hypothesized during on-line word identification.

Initially, the data from the Harvard and Haskins material sets were analyzed separately for gating-direction effects and then planned comparisons between the material sets were carried out. The data for the words in each sentence position were pooled across the eight test sentences. This was done to test for serial-order effects, that is, the possibility that recognizing words early in the sentence might influence the identification of words occurring later in the same sentence (Cole & Jakimik, 1980). The results for the analyses of identification points and the incorrect word responses are examined separately below.

Identification Point Results—The identification point data are shown in Figure 2 for words in meaningful and anomalous sentences (and their isolated controls) in the left- and right-hand panels, respectively. In each panel, identification points are shown for words in each sentence position for both gating directions; mean identification points averaged over all sentence positions are shown in the far right of each panel. The means of the measured

³One incorrect response after at least two consecutive correct identification trials that was corrected on following trials was allowed in our scoring procedure. Our identification points were operationally defined and differed from both Grosjean’s “isolation points” (1980) and Ottevanger’s “recognition points” (1980, 1981). Grosjean’s phrase “identification point” is an empirical term used within the theoretical framework of cohort theory (see Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1980; Tyler & Marslen-Wilson, 1982b). In this usage, the concept of an identification point differentiates between the reduction of the cohort set to a single word and subjects’ confidence of their conscious identification responses. Ottevanger’s term “recognition point” also adheres to the theoretical definition of the cohort framework, that is, “the phoneme going from left to right in the word that distinguishes that word from all others beginning with the same sound sequence” (Ottevanger, 1980, p. 85; See also Tyler & Marslen-Wilson, 1982b). In the present study, identification points correspond to the amount of signal duration required for consistently correct identification of a word by listeners. Thus, our use of the term “identification point” is operational and is not based on theoretical constructs associated with any particular theory.

physical durations of words at each sentence position are also included as a baseline for comparison.

The raw identification points were converted into proportions of the mean word duration in each sentence position to compensate for differences in duration as a function of syntactic structure.⁴ Statistical analyses were then carried out on arcsin transformations of the proportions. Analyses of variance by subject and treatment were performed. Gating direction and sentence position (or subjects) were treated as fixed and random factors, respectively, and min F' statistics were calculated (Clark, 1973). Unless otherwise stated, all significance levels are less than $p = .01$.

Four major findings were obtained for words in the meaningful sentence contexts. First, backward-gated words required greater signal duration for identification than forward-gated words. This was true for words in a sentence context, $\min F'(1,78) = 6.56$, and for their isolated controls, $\min F'(1,90) = 7.30$, respectively. Second, word identification required more signal duration in isolation than in meaningful sentences, $t(78) = 6.22$: The mean difference was 96 milliseconds. Third, the presence of normal sentence context decreased the listeners' dependence on the acoustic-phonetic information in the signal compared to isolated word identification. The amount of signal duration required to identify the isolated words excised from the Haskins anomalous sentences correlated positively with their measured durations, $r(38) = .89$. No such relationship was observed for the identification points and durations of the words excised from the meaningful sentences, $r(38) = .24$, $p > .10$. Fourth, no main effect of sentence position was observed, $F(4,44) = 1.67$, $p > .1$. This result suggests that the words that occurred earlier in a sentence apparently conveyed no predictive information that enabled listeners to identify the following words at shorter gate durations.

The identification point data for words from the anomalous (Haskins) materials can be summarized as follows: First, forward-gated words in isolation required less signal for identification than backward-gated words, $\min F'(1,80) = 13.84$, but this was not true in the anomalous context, $\min F'(1,56) = 1.26$, $p > .10$. No main effect of anomalous context compared to the isolated control condition was observed. Second, the presence of anomalous sentence context did not reduce the listeners' dependence on the acoustic-phonetic information in the signal: Identification points correlated positively with the measured word duration for both words in anomalous contexts and their isolated controls, $r(30) = .92$, and $r(30) = .76$, respectively. Third, no sentence position effect was found in the presence of anomalous contexts: Neither the subject-wise nor the treatment-wise test was significant ($F_s(3,129) < 1.0$, $p > .44$ and $F_s(3,29) = 1.43$, $p > .25$).⁵ It appears that the anomalous sentence context somehow inhibits the normal reliance on acoustic-phonetic information for word identification.

The identification data for words in both meaningful and anomalous sentences are shown in Figure 3 as proportions of the mean word durations in each sentence position. No differences were observed for the isolated words: Words excised from the Harvard and Haskins sentences were identified with .83 and .81 of the mean word duration, respectively.

⁴Words in sentence-final or phrase-final positions are generally acoustically longer than words in other sentence positions. The use of proportions to normalize for these effects enabled us to compare identification points across sentence positions, context types, and different durations.

⁵The variation in identification points for words in the anomalous context according to sentence position shown in Figure 2 was not observed when the data from the anomalous sentences were displayed as proportions in Figure 3. The shape of the curve in Figure 2 reflects the fact that the form class and length of the target words in the Haskins anomalous sentences were confounded with sentence position. All the sentences had the same surface structure (i.e., Determiner-Adjective-Noun-Verb-Determiner-Noun). The words in the second and fourth positions in the sentences were always nouns and were physically longer in duration than words in all other sentence positions.

However, in the anomalous sentence context, .72 of the mean word duration was needed for identification, while only .56 was necessary to identify the words in the meaningful sentences, $t(82) = 3.31$. Of the four context conditions, only in the anomalous sentence context was no advantage for word-initial acoustic–phonetic information over word-final acoustic–phonetic information observed. This result suggests that an anomalous sentence context prevents the listener from using word-initial information in the same way as it is used in normal sentence environments. In this case, lexical access processes are misdirected and listeners require more signal duration to identify words.

Analysis of the Response Distributions—We were also interested in the distribution and structural organization of the incorrect word candidates generated by listeners before they correctly identified the target words in the different context conditions. The number of different incorrect word responses proposed by at least one subject was examined as a measure of response output. Analyses of variance with gating direction and sentence position as factors and sentences as repeated measures were performed on these output measures.

The mean number of different word responses in each sentence position in the meaningful and anomalous sentences (excluding correct identification responses) are shown in Figure 4. For the words in meaningful sentences, a gating-direction effect was found, $F(1,14) = 4.94$, $p < .05$: Word-final acoustic–phonetic information yielded more word candidate responses than word-initial acoustic–phonetic information. This was not observed for words in the anomalous sentences, $F(1,14) < 1.0$, $p > .56$. Sentence position effects were found for both context types, $F(4,56) = 5.75$ and $F(3,42) = 8.68$. However, the nature of these position effects differs substantially. In the meaningful sentences, fewer incorrect responses were proposed for words that occurred later in a sentence than for words that occurred earlier in the sentence. This is shown in the decreasing slope of the two curves in the left-hand panel of Figure 4. In contrast, in the anomalous sentences, more word candidates were generated for words in the second and fourth sentence positions than in the first and third positions. These data parallel the variations in the identification point data for the anomalous context condition (see Figure 2 and footnote 5). In fact, a significant correlation was observed between identification points and the number of incorrect word responses for the anomalous context, $r(30) = .96$. (This finding was not observed for the identification points and number of incorrect word candidates in the meaningful context, $r(38) = .26$, $p > .10$.)

Finally, more incorrect lexical candidates were proposed in the semantically anomalous sentences than in the meaningful sentences, $t(70) = 5.41$. This result occurred despite the fact that the sentence frame was always fixed for the anomalous sentences, which we expected to aid identification. Analyses of the number of word candidates support our interpretation that the presence of an anomalous context interferes with the normal integration of acoustic–phonetic and contextual information that leads to word identification in sentences.

To examine the structure of the word candidate responses in greater detail, analyses of the hypothesized sources of knowledge underlying the proposed word candidates were carried out for each subject's response protocol. Each incorrect word response was categorized as originating from one of three possible sources: (1) acoustic–phonetic analysis of the signal; (2) “syntactic” contextual information; and (3) “other” sources (nonwords, words from an inappropriate form class, or intrusions). Incorrect word candidates were classified as belonging to only one of these three categories. In this scoring procedure, preference was given to acoustic–phonetic information as a knowledge source, so that the remaining two categories contained no candidates that were phonetically similar to the target word.⁶ Thus, we selected a conservative measure of the nonsensory contributions to the set of words

hypothesized before a word was identified correctly. Within the “acoustic” category, our scoring procedure prevented discrimination of contextually appropriate and inappropriate word candidates.

We assumed that in the meaningful sentences, normal pragmatic, semantic, and syntactic constraints were used by listeners to identify the target words. In contrast, we assumed that no normal pragmatic or semantic relations could be derived from the anomalous sentences. In fact, whatever semantic cues might be generated would most likely be incompatible with the normal syntactic cues in this condition. The criterion we adopted for scoring membership in the “syntactic” knowledge category was based solely on appropriate form class (in the absence of correct acoustic–phonetic information) for both meaningful and semantically anomalous sentences. We note that semantic and syntactic sources of knowledge were not differentiated within the “syntactic” knowledge category. Although this knowledge source could not be used in the isolated word identification task, response distributions in those conditions were nevertheless scored for this category also, simply as a baseline control measure.

Finally, the “other” category of incorrect responses contained primarily response intrusions from other sentences or from other serial positions in the same test sentence. Also included in this category were phonemically dissimilar nonwords and words from inappropriate form classes. Since the “other” word responses constituted less than 4% of the total incorrect responses, they were omitted from statistical analyses performed on the response distributions.

Table 1 shows the distribution of incorrect responses over all sentence positions and gate durations. Data from the three categories of incorrect word candidates, and blank responses are given as proportions of the total number of incorrect responses. In all conditions, the majority of incorrect responses were blanks. However, our treatment addresses the distributions of acoustic and syntactic word candidates, since they can be related to underlying sources of knowledge available to the listener, whereas this may be difficult for other kinds of errorful responses.

Two major results were observed in the analyses of the response distributions. First, more word candidates were based on acoustic than on syntactic information. This was true for Harvard words both in meaningful contexts, $F(1,28) = 9.83$, and in isolation, $F(1,28) = 286.68$, and for Haskins words both in their anomalous contexts, $F(1,28) = 225.59$, and in isolation, $F(1,28) = 428.11$, respectively. Second, the role of nonacoustic, top-down knowledge sources in the response distribution increased in both meaningful and anomalous contexts compared to the isolated word conditions: For both the Harvard and Haskins words, more syntactically based word responses were made in the sentence context than in isolation, $F(1,28) = 5.75$, $p < .02$ and $F(1,28) = 13.63$, respectively. This result suggests that there are no syntactic constraints operating in the isolated control conditions. The observed facilitation from the anomalous context suggests that subjects may have taken advantage of the sentence frame similarity between all the Haskins sentences. Thus, semantically anomalous sentence context appears to inhibit normal efficient use of knowledge sources to identify words correctly, but in the process, the context causes more incorrect candidates, particularly from nonacoustic sources, to be considered. When incorrect word candidates

⁶In the absence of any standardized criteria for degree of phonetic overlap, the following guidelines were adopted by the first author and a research assistant (N.C.). Similarity between the initial (or final) phoneme of the response and the target word received greatest weighting. Words whose initial phoneme differed only in its voicing feature from that of the target word, according to the Chomsky and Halle feature system (1968), were included in the category as well. Finally, responses which retained the vowel and at least one other phoneme from the target word in the correct sequence were also considered to be based primarily on acoustic–phonetic information contained in the signal.

were generated, they were based, in large part, on the acoustic–phonetic information present in the speech signal. However, the results for the syntactically based candidates stretch the spirit of “bottom-up priority,” whereby the set of possible word candidates is initially defined exclusively by sensory input (Tyler & Marslen-Wilson, 1982b).

When the data for the words in meaningful sentences and in isolation are compared, the effects of the meaningful sentence context are readily apparent. The relative contribution of acoustic information did not differ in isolation and in meaningful context, $F(1,28) = 1.18$, $p > .28$. In contrast, more syntactically based word candidates were generated in the sentence context than in isolation, $F(1,28) = 5.75$, $p < .02$. A significant Gating-Direction $\times$ Sentence Context interaction was observed for the syntactically based responses, $F(1,28) = 7.77$. When only word-final acoustic–phonetic information was present in the meaningful sentences, the number of responses based on correct syntactic knowledge increased.

It appears that the presence of meaningful sentence contexts operates to expand the role of syntactic information: In particular, the presence of meaningful context co-occurring with the absence of word-initial acoustic–phonetic information appears to increase the reliance on syntactic contextual information in hypothesizing word candidates. Thus, subjects are able to identify words from word-final information in sentences, but they do this in ways that are qualitatively and quantitatively different from identification based on word-initial information.

Our results from identification points and response distributions support three broad claims. First, the acoustic–phonetic information contained in the beginnings of words is more useful to listeners compared to the information contained in the ends of words. However, word identification is possible without information from the beginnings of words. This result calls into question the obligatory status given to word-initial information in a number of recent accounts of spoken word identification (Cole & Jakimik, 1980; Forster, 1976, 1979; Marslen-Wilson & Welsh, 1978; Tyler & Marslen-Wilson, 1982a, 1982b). Second, meaningful sentence contexts support faster, more efficient, and qualitatively different identification processes than semantically anomalous sentence contexts or the presentation of words in isolation. Third, we have demonstrated a reliable relationship between the structure of the set of incorrect word responses and the final product of the word identification process. When less dependence on acoustic–phonetic information exists as in most normal sentence processing situations, identification occurs faster and more accurately. In our study, listeners identified forward-gated words in meaningful sentences with the shortest gates and the fewest incorrect lexical responses. Thus, both accuracy and temporal measures of perceptual processing allowed by our gating procedure support the claim that spoken word identification in sentences is most efficient when word–initial information and normal sentential constraints are present.

Experiment 2

The aim of Experiment 2 was threefold: first, to replicate the effects observed in the first study; second, to study the changes in the distribution of word candidates over increasing signal durations; and third, to investigate the effects specific to the successive presentation procedure used in Experiment 1. In this experiment, we made one additional assumption, namely, that the sources of knowledge used in lexical access and word identification would be differentially informative at various points in the time course of the word identification process. Thus, we expected some variation in the balance of lexical responses based on different knowledge sources as a function of the time course of the word identification process. These changes should be reflected in the set of word candidates generated at different gate durations. These assumptions yield several predictions that can be used to test

the autonomous character of lexical access processes as they have been specified in cohort theory (Marslen-Wilson & Welsh, 1978; Tyler & Marslen-Wilson, 1982b). According to the cohort theory, the original set of word candidates is activated solely by the acoustic–phonetic information in approximately the first 175–200 milliseconds of the speech signal at the beginnings of words. By this view, decreases in the set of hypothesized words during the time course of word identification are seen as the consequence of both the sensory input *and* the top-down syntactic and semantic knowledge available from the context. A word is identified or recognized when all but one lexical candidate are deactivated by the interaction of these two knowledge sources (see Grosjean, 1980; Tyler & Marslen-Wilson, 1982b).

In the present experiment, we predicted that at very short gate durations, when minimal acoustic–phonetic segmental information is available in the speech signal, a larger proportion of word candidates should be based on other knowledge sources than at longer gate durations. By this prediction, interactive processes can provide input to the set of lexical candidates generated before a word is consciously identified. If, on the other hand, semantic and syntactic knowledge can only be used to eliminate incorrect (acoustic–phonetic based) word candidates, as Marslen-Wilson and his colleagues claim, then any nonacoustic syntactic word candidates occurring in the response distribution should simply represent random noise. To this end, we examined the time course or growth of word candidates hypothesized by listeners in terms of changes in the distribution of word candidates over successive gate durations.

Several methodological questions were also addressed in the present experiment. First, the procedure of successive gated sentence trials (used in Experiment 1 and by Grosjean, 1980), each with the presentation of greater and greater signal durations, may have influenced subjects' word identification responses artifactually. Repeated presentations of the same signal with short gate durations may have led to facilitation in terms of the amount of signal duration required for word identification. On the other hand, subjects may have developed specialized response strategies during successive presentations of a test sentence. Having available their responses from earlier trials on their answer sheets may have influenced subjects' subsequent responses on later presentations of the same test signal. Conceivably, subjects may have been reluctant to change some word candidates, even when additional acoustic–phonetic information was present. To determine the validity of the procedure used in Experiment 1, several procedural and design changes were made. First, in a within-subject design, each subject heard each test sentence only once and every subject heard both meaningful and anomalous sentence contexts mixed in both forward and backward gating conditions. Second, we were interested in the effects of the syntactic information contained in the printed sentence frames on subject answer sheets in Experiment 1. The data from the sentence context and isolated word conditions in Experiment 1 suggest that syntactic knowledge plays only a minimal role in spoken word identification processes. Subjects may have used general linguistic knowledge, as opposed to specific contextual knowledge gained from a bottom-up parsing analysis of each stimulus input (Garrett, 1978). If this was indeed the case, then the function word sentences frames, for example, "The _____ in the _____," would not be instrumental in providing subjects with syntactic information specific to each test sentence in Experiment 1. In the present study, therefore, subjects had no visual information about the semantic or syntactic structure of the sentences: Subjects simply wrote down whatever words they heard after each sentence presentation.

Method

Subjects—The subjects were 64 students drawn from the same pool as those for Experiment 1. None of the subjects participated in the first experiment.

Materials—A subset of the materials used in Experiment 1 was chosen to include 16 gating conditions for each of eight meaningful and eight semantically anomalous sentences. In the 0-millisecond gate duration, every content word was entirely replaced with envelope-shaped noise. In the “Full” condition the intact spoken sentence was presented. The 50, 100, 150, 200, 250, 300, and 350-millisecond gates of signal duration were employed in both forward-gated and backward-gated conditions. Each test sentence was presented only once during a trial, in contrast to the repeated-presentation format of test sentences in Experiment 1.

Design and Procedure—In a latin-square design, 16 groups of four subjects listened to each of the 16 original sentences in a different gating condition. Each group was given two practice sentences at the beginning of the experimental session. Subjects were instructed to listen and try to identify the words in each sentence as completely as possible on blank answer sheets. Subjects were not informed of the number of words in each sentence or the differences between meaningful and anomalous sentence contexts. A PDP-11/34 mini-computer controlled the order and presentation of the stimuli. Every group heard a different random order of their particular gating conditions of the eight meaningful sentences followed by eight anomalous sentences. A trial began with a 500-millisecond cue light, which was followed by a 1-second pause. Subjects were then presented with the gated sentence, which was output at 77 dB SPL through their headphones. After writing their responses down, subjects pressed a “Ready” button on a response console. When all subjects in a group had responded, the next trial was automatically initiated. Every subject heard each of the 16 sentences once, each in a different gating duration-by-direction condition. Experimental sessions lasted approximately 20 minutes.

Three variables were examined: gating direction (forward vs backward), context type (meaningful vs anomalous sentences), and for comparisons with Experiment 1, presentation type (single vs repeated presentations). For each of these variables, we were interested in both the amount of signal duration required to identify the words in the sentences and the nature of the distribution of incorrect responses.

Results and Discussion

For each gating duration, the probability of correct identification was calculated and incorrect word responses were recorded and categorized as in Experiment 1. The correct identification data and the analysis of incorrect word candidates are discussed in separate sections below. Unless otherwise indicated, all reported statistical tests are significant at the $p = .01$ level.

Identification Data—Figure 5 shows the probability of correct identification of words in the meaningful and syntactic sentences as a function of signal duration. χ^2 tests were used to determine the differences between the psychometric curves for gating direction and context. For each test, two rows (forward and backward gating direction, or meaningful and anomalous context) and nine columns of signal durations were used. Best-fitting logistic functions using the method of least squares were calculated for each identification curve and were used to obtain the .50 probability of correct identification for each condition.

As in Experiment 1, an advantage of word-initial acoustic–phonetic information over word-final acoustic–phonetic information was observed across both the meaningful and anomalous sentence contexts, with a single presentation of words in a gated sentence. The probability of correct identification was greater for forward-gated words than for backward-gated words in both the meaningful context condition, $\chi^2(8) = 24.84$, and the anomalous context condition, $\chi^2(8) = 24.78$, respectively. At .50 probability of identification, the identification threshold, the word-initial advantage was 46 milliseconds in meaningful

sentences and 39 milliseconds in anomalous sentences. The advantage for word-initial information was significant within the range of signal durations between 100 and 300 milliseconds. These results replicate our earlier findings with the meaningful sentence context, and demonstrate that the repeated presentation format used in Experiment 1 was valid for normal sentence conditions.

In Figure 5, the forward and backward-gated psychometric curves for the anomalous sentences (lower panel) are shifted to the right compared to the corresponding curves for the meaningful sentences (top panel). For forward-gated words, a 31-millisecond advantage at the .50 level of identification accuracy was found for words in the meaningful sentences, $\chi^2(8) = 34.94$; the magnitude of the corresponding difference for backward-gated words was 24 milliseconds, $\chi^2(8) = 23.59$. That is, word identification was more accurate in the meaningful contexts than in the anomalous contexts, when the speech signal was equated in terms of the amount of available acoustic–phonetic information in each word.

One methodological concern of this study was a comparison of the single-presentation procedure with the successive presentation method in Experiment 1. In order to compare data on the amount of signal duration required with single sentence presentation to the amount required with successive sentences in Experiment 1, the identification point data from that study were converted to probabilities of correct identification for each condition.⁷ The psychometric curves for the meaningful and anomalous sentences from Experiments 1 and 2 are shown in the top and bottom panels of Figure 6, respectively. (The single presentation data from the present study that were displayed in Figure 5 are repeated for comparison.)

Presentation format interacted with context: For the meaningful sentences, no presentation effect was observed, $\chi^2(8) = 7.64$, $p > .45$ and $\chi^2(8) = 14.80$, $p > .08$, for forward-gated and backward-gated words, respectively. Surprisingly, the identification thresholds we calculated in the two experiments for forward-gated words in the meaningful sentences were almost identical—161 and 166 milliseconds, respectively. Similar findings have been reported recently by Cotton and Grosjean (1984). For both presentation methods, nevertheless, forward-gated words in meaningful sentences were identified with less signal duration than backward-gated words. Thus, the gating-direction effect in meaningful context is robust across procedural differences of single versus successive sentence presentations and across different experimental designs (within subjects vs between subjects). We therefore conclude that the use of the successive sentence presentation method in Experiment 1 did *not* artifactually facilitate the identification of words in the meaningful sentences (see also Cotton & Grosjean, 1984).

In contrast, however, for the words in the anomalous sentence contexts a dramatic reversal was observed: Presentation format did affect the amount of signal duration required for word identification. With single presentations, *less* signal duration was required for subjects to correctly identify words in anomalous sentences. Identification of forward-gated words was actually inhibited when the words were presented repeatedly in anomalous sentence contexts as in Experiment 1, $\chi^2(6) = 38.96$. Conflicting top-down semantic and syntactic information may have accumulated with successive repetitions of the same anomalous context to prevent efficient and rapid use of word-initial acoustic–phonetic information. As a result, for words in anomalous sentences, the probability of identification in the successive presentation

⁷To compute these values, the data were rescored so that each identification point in Experiment 1 contributed to the identification probability curve at every gate duration shorter than itself. When calculated in this way, each subject in Experiment 1 contributed to many gate durations, corresponding to successful identification of a word on consecutive sentence presentations but different groups of subjects contributed to each direction and context condition curve. In contrast, the points on each identification curve in the present experiment reflect independent data points with the same subjects contributing to all four curves.

forward-gated condition (Experiment 1) was, in fact, increased artifactually relative to the single presentation condition (Experiment 2).

In short, the effects of presenting successive sentence trials were minimal in the condition that best simulates natural speech processing situations, that is, the forward-gated meaningful sentences. However, repeated presentations of conflicting semantic and syntactic information in the anomalous sentences caused subjects in Experiment 1 to adopt a processing strategy that required them to use more signal duration in order to recognize words than in the single-presentation procedure of the present experiment.

The identification data obtained in the present study replicated the major findings of Experiment 1. On a single presentation, word-initial acoustic–phonetic information was more informative than word-final acoustic–phonetic information. This advantage of word-initial signal occurred for both meaningful and semantically anomalous sentence contexts in the present single-presentation procedure although it was obscured by the successive presentation condition in Experiment 1 for the semantically anomalous sentence contexts. The failure to observe facilitation due to forward gating in the anomalous sentence conditions of Experiment 1 suggests a process involving the slow accumulation of conflicting top-down constraints in identification of words in the anomalous sentences.

Analysis of the Response Distributions—The incorrect response distributions in the single-presentation conditions are given in Table 2. The data are displayed as proportions of total errorful responses for meaningful and anomalous contexts in the left- and right-hand panels, respectively.

The complex pattern of word response distribution results obtained for Experiment 1 was replicated in the present study. In both experiments, more incorrect word candidates were generated for words in semantically anomalous sentence contexts than for words in meaningful sentence contexts. Subjects in both experiments also generated more incorrect responses based on acoustic–phonetic information in the anomalous context conditions. Moreover, in meaningful contexts, more incorrect word candidates were based only on appropriate syntactic constraints than in the anomalous context conditions. When only word-final acoustic–phonetic information was presented, the number of syntactically based word responses was greater than when word-initial acoustic–phonetic information was available for the words in the meaningful sentences. Thus, the present findings demonstrate that the observed distributions of word candidates found in Experiment 1 were not in any way artifacts of the successive presentation procedure or that the meaningful and semantically anomalous sentences were presented in a blocked format to separate groups of subjects. More importantly, however, a similar trade-off of sources of knowledge between acoustic–phonetic information and syntactic information in generating word candidate responses was observed for the meaningful sentences across the two experimental designs and procedures.

The primary theoretical goal of Experiment 2 was the analysis of the development of the distribution of word candidates with increasing signal duration, what we have called cohort “growth functions.” This analysis may be used to characterize the relative time course of top-down and bottom-up contributions to specification of the set of potential lexical candidates. In Figure 7, growth functions of the response distributions are shown for the meaningful and anomalous sentence conditions, in the left- and right-hand panels, respectively. These are broken down by correct responses, acoustic–phonetic, and syntactically based incorrect word responses. Forward-gated and backward-gated conditions are shown in the top and bottom panels, respectively. Two classes of responses are not shown: “other” responses, which accounted for less than 2% of the total errors, and blank responses. To fill out the distribution at each gate duration, the bulk of missing data comes

from blank responses, which generally were very numerous at short gates and decreased rapidly as more acoustic–phonetic information was available.

Three findings may be seen in Figure 7: First, as shown previously in Figure 5, in each condition the proportion of correct responses increases with longer signal durations. Second, the proportion of incorrect lexical responses was quite small overall when compared to the proportion of correct responses. Nevertheless, the number of incorrect word responses peaked at short signal durations, well before the identification threshold at .50 accuracy was reached, and then declined at longer gate durations. Third, considered over all gate durations, a relatively larger proportion of incorrect word responses was based on acoustic–phonetic information in the semantically anomalous contexts than in the meaningful contexts, suggesting that subjects had to rely more on this source of knowledge than on other sources in identifying words in these unusual contexts.

One feature of the word candidate growth functions is particularly noteworthy: the relative peaks in the growth functions for the two categories of incorrect responses. For the meaningful contexts, the greatest number of word candidates based on compatible word form class but incorrect information occurred at the 100-millisecond duration. In contrast, for both meaningful and anomalous sentence contexts, the acoustic–phonetic controlled response distribution reached its peak at a longer gate duration and, unlike the syntactically controlled responses, maintained its proportion of the response distribution over approximately the next 100 milliseconds of signal duration. This can be seen in Figure 7 by comparing the peaks of the distributions of responses based on syntactic and acoustic–phonetic information, represented by squares and triangles, respectively, in each panel.

In the backward-gated meaningful context conditions, the word candidates generated on the basis of syntactic knowledge dominate the distribution at the 0, 50, and 100-millisecond gate durations. By 150 milliseconds, both correct responses and acoustic–phonetic responses surpassed the contribution of purely syntactically based lexical candidates. Thus, when only word-final acoustic–phonetic information was available, the syntactic cues in the meaningful sentences allowed for an early contribution of syntactic knowledge to the set of potential word candidates.

This pattern of results suggests that for meaningful contexts, top-down syntactic knowledge may be used to generate word responses fairly early in the identification process, even if the word candidates are incompatible with the actual acoustic–phonetic input. Syntactic knowledge therefore appears to contribute to the set of hypothesized word candidates even before enough acoustic–phonetic information is available to initiate lexical access from bottom-up sources. However, when at least 150 milliseconds of acoustic–phonetic information is available in the speech signal, only words that are also acoustic–phonetically compatible with the input are maintained as potential candidates in the response distribution. At this point, the relative contribution of syntactically controlled incorrect lexical candidates appears to decrease while the pool of acoustically-based incorrect lexical candidates increases.

The top and bottom right-hand panels of Figure 7 show the growth functions of the response distributions for the forward-gated and backward-gated words in the anomalous sentence context, respectively. Again, the number of incorrect word candidates based solely on acoustic–phonetic information begins to decline at shorter signal durations for forward-gated words than for backward-gated words. This relation between the peaks of the growth functions of the acoustic–phonetic word candidates corresponds to the gating direction difference observed in the identification thresholds in Figure 5. More responses based on appropriate syntactic knowledge occurred on trials in which between 50 and 150

milliseconds of signal duration was presented. Thus, a small number of syntactically appropriate word candidates were generated even in the semantically anomalous contexts when negligible acoustic–phonetic information was available at the shortest signal durations.

Our analyses of the growth functions of the response distribution suggest that subjects used general syntactic knowledge to generate word responses in gating conditions in which the signal duration was less than 150 milliseconds. With longer signal durations, the number of purely syntactically based word responses decreased to a small proportion of the total response candidates hypothesized in both meaningful and anomalous contexts. When compatible semantic information accompanied the available syntactic cues, as in the meaningful sentences, the proportion of syntactically controlled responses increased. Thus, in meaningful sentences, subjects employed both bottom-up *and* top-down information to hypothesize word candidates when less than 150 milliseconds of the signal was available. When more reliable acoustic–phonetic information about the word is provided at longer gate durations, the contribution of syntactic and semantic knowledge appears to be less prominent. Like Grosjean (1980), we found that nonacoustic sources of knowledge contribute to incorrect responses at short gates and that the pool of word candidates is not largest at short gates, but instead, peaks just before enough bottom-up information is available for accurate word identification.

In summary, the results of Experiment 2, using a different presentation procedure and a different experimental design, replicated the major findings obtained in Experiment 1. In addition, the present study identified some of the temporal characteristics of the interaction between bottom-up and top-down knowledge sources in word identification. In both presentation methods, we found an advantage of word-initial acoustic–phonetic information over word-final acoustic–phonetic information in identifying words in meaningful sentences. The amount of signal duration required for correct word identification and an analysis of the number of incorrect word responses supported these observations. Further analysis of word candidates generated at increased signal durations revealed that the distributions of hypothesized words were quite sensitive to differences in the use of the various knowledge sources available in both the meaningful and semantically anomalous sentences.

General Discussion

The results from our two experiments demonstrate that words can be identified in sentences without word-initial acoustic–phonetic information. However, the distribution of the hypothesized word candidates appears to be sensitive to the presence of normal sentential semantic and syntactic constraints and generally appears to be dominated by bottom-up acoustic–phonetic information. Our data suggest that in normal processing of fluent speech, the acoustic–phonetic information present at the beginnings of words is a major source of information used for word recognition, lexical access, and the subsequent word identification process. While the processing system also uses nonacoustic sources of knowledge, when normal continuous speech processing conditions are simulated, word-initial acoustic–phonetic information appears to control the distribution and selection of potential word candidates when one condition is met, namely, that at least 150 milliseconds of the beginning of a word have been presented. This estimate, which is derived from our specific stimuli, a phonetically balanced subset of content words in English, is similar to Marslen-Wilson's estimate of two word-initial phonemes (Marslen-Wilson, 1980) or roughly the duration of a syllable. Thus, words can be identified from relatively brief segments of the acoustic waveform in meaningful sentences through the contribution of semantic and syntactic sentential constraints. Our analysis of the errors indicated that word

candidates based on nonacoustic sources were frequently hypothesized when only small segments of the acoustic–phonetic input were present in the speech signal.

The present results have two important implications for understanding context effects in speech perception and spoken word identification. First, in both experiments we found that word identification in semantically anomalous sentences did *not* resemble the corresponding processes observed in meaningful sentences. Previous studies have ordered contextual manipulations along a continuum (e.g., Grosjean, 1980; Marslen-Wilson & Tyler, 1980; Miller et al., 1951). Our results suggest an alternative analysis of context manipulations. The similarities in perception between isolated words and words in anomalous sentences in Experiment 1 suggest that in these impoverished contexts, words may have been processed more as though they occurred in randomized lists without any internal structural organization than as “syntactically normal” sentences. We suggest that there are important qualitative and quantitative differences between meaningful and anomalous contexts. It seems plausible to us that the inhibition observed in the anomalous contexts indicates the operation of a nonautonomous or interactive syntactic processor and/or integration stage of comprehension (Cairns, 1982). Our data, therefore, support proposals of the parallel and interactive operation of syntactic and lexical processing of words in spoken meaningful sentences where semantic constraints constitute an obligatory source of information for the on-going operation of normal lexical access and word identification processes. When semantic information is experimentally removed or selectively manipulated in aberrant ways, the listener responds with strategies to solve the identification task that do not necessarily reflect the normal interplay of knowledge sources in spoken language processing.

Second, the present results also address the issue of autonomy in language processing, specifically as it relates to lexical access and word identification. Our findings provide empirical support for Marslen-Wilson’s “principle of bottom-up priority” in a variety of conditions, but only within the framework of an interactive word identification process. While acoustic–phonetic information in the speech signal appears to be the primary source of information used by listeners to generate lexical candidates accessed from long-term memory in the first stage of word identification, semantic and syntactic information provided by the sentence context also enables syntactically compatible words that are not based on the sensory input to be activated and entered into the pool of hypothesized word candidates. As the acoustic–phonetic determination of a word begins to emerge, fewer and fewer word candidates need to be considered by the listener. Thus, listeners appear to use all the available information in both stages of spoken word identification, weighting, if only momentarily, the most reliable knowledge source most. Before acoustic information has accumulated to chunks of approximately 150 milliseconds, syntactic knowledge does appear to play a role in constraining the set of potential lexical candidates. At this point in time, acoustic–phonetic information gains prominence while both top-down and bottom-up sources continue to eliminate incorrect candidates from the hypothesized set of words.

The presence of compatible semantic and syntactic information therefore appears to be an obligatory component of normal word identification in meaningful sentences. The relative balance among these various sources of information provides an extremely efficient system to support spoken language comprehension over a wide range of listening conditions in which the physical signal may be degraded or impoverished. The operation of these higher-level sources of knowledge at both perceptual and post-perceptual stages therefore allows listeners to perceive and understand spoken language even under adverse conditions.

Acknowledgments

We wish to thank Michael Studdert-Kennedy for insightful discussions; Nancy Cooney for help in collecting and analyzing the data; and Tim Feustel, Marcel Just, Paul Luce, Howie Nusbaum, Amanda Walley, and two anonymous reviewers for their comments on an earlier version of the manuscript. This research was supported, in part, by NIMH Grant MH-24027 and NINCDS Grant NS-12179 to Indiana University in Bloomington.

References

- Bagley WC. The apperception of the spoken sentence: A study in the psychology of language. *American Journal of Psychology*. 1900; 12:80–130.
- Bruner JS, O'Dowd D. A note on the informativeness of parts of words. *Language and Speech*. 1958; 1:98–101.
- Cairns, HS. Unpublished manuscript. 1982. Autonomous theories of the language processor: Evidence from the effects of context on sentence comprehension.
- Chomsky, N.; Halle, M. *The sound pattern of English*. New York: Harper & Row; 1968.
- Clark HH. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*. 1973; 12:335–359.
- Cole, RA.; Jakimik, JA. A model of speech perception. In: Cole, RA., editor. *Perception and production of fluent speech*. Hillsdale, N.J: Erlbaum; 1980.
- Cole RA, Rudnick A. What's new in speech perception? The research and ideas of William Chandler Bagley, 1874–1946. *Psychological Review*. 1983; 90:94–101. [PubMed: 6342015]
- Cotton S, Grosjean F. The gating paradigm: A comparison of successive and individual presentation formats. *Perception and Psychophysics*. 1984; 35:41–48. [PubMed: 6709473]
- Egan JP. Articulation testing methods. *Laryngoscope*. 1948; 58:955–991. [PubMed: 18887435]
- Forster, KI. Accessing the mental lexicon. In: Walker, E.; Wales, R., editors. *New approaches to language mechanisms*. Amsterdam: North-Holland; 1976.
- Forster, KI. Levels of processing and the structure of the language processor. In: Cooper, WE.; Walker, ECT., editors. *Sentence processing: Psycholinguistic studies presented to Merrill Garrett*. Hillsdale, N.J: Erlbaum; 1979.
- Garrett, MF. Word and sentence perception. In: Held, R.; Leibowitz, HW.; Teuber, H-L., editors. *Handbook of sensory physiology: Vol. VIII. Perception*. Berlin: Springer-Verlag; 1978.
- Grosjean F. Spoken word recognition processes and the gating paradigm. *Perception and Psychophysics*. 1980; 28:267–283. [PubMed: 7465310]
- Horii Y, House AS, Hughes GW. A masking noise with speech envelope characteristics for studying intelligibility. *Journal of the Acoustical Society of America*. 1971; 49:1849–1856. [PubMed: 5125732]
- Marslen-Wilson, WD. Speech understanding as a psychological process. In: Simon, JC., editor. *Spoken language generation and understanding*. Dordrecht: Reidel; 1980.
- Marslen-Wilson WD, Tyler LK. The temporal structure of spoken language understanding. *Cognition*. 1980; 8:1–71. [PubMed: 7363578]
- Marslen-Wilson WD, Welsh A. Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*. 1978; 10:29–63.
- Miller GA, Heise GA, Lichten W. The intelligibility of speech as a function of the context of the test materials. *Journal of Experimental Psychology*. 1951; 41:329–335. [PubMed: 14861384]
- Miller GA, Isard S. Some perceptual consequences of linguistic rules. *Journal of Verbal Learning and Verbal Behavior*. 1963; 2:217–228.
- Morton J. Interaction of information in word recognition. *Psychological Review*. 1969; 76:165–178.
- Morton, J. Facilitation in word recognition: Experiments causing change in the logogen model. In: Kolers, PA.; Wrolstal, ME.; Bouma, H., editors. *Processing of visible language I*. New York: Plenum; 1979.
- Nooteboom SD. Lexical retrieval from fragments of spoken words: Beginnings versus endings. *Journal of Phonetics*. 1981; 9:407–424.

- Norris D. Autonomous processes in comprehension: A reply to Marslen-Wilson and Tyler. *Cognition*. 1982; 11:97–101. [PubMed: 7198955]
- Norris, D. Unpublished manuscript. 1982b. Word recognition: Context effects without priming.
- Nye, PW.; Gaitenby, J. Haskins Laboratories Status Report on Speech Research. 1974. The intelligibility of synthetic monosyllable words in short, syntactically normal sentences; p. 169-190.
- Ohman S. Perception of segments of VCCV utterances. *Journal of the Acoustical Society of America*. 1966; 40:979–988. [PubMed: 5956022]
- Ottevanger IB. Detection of mispronunciations in relation to the point where a word can be identified. *Progress Report Institute of Phonetics University of Utrecht*. 1980; 5:84–93.
- Ottevanger IB. The function of the recognition point in the perception of isolated mispronounced words. *Progress Report Institute of Phonetics University of Utrecht*. 1981; 6:54–69.
- Pisoni DB. Some current theoretical issues in speech perception. *Cognition*. 1981; 10:249–259. [PubMed: 7198545]
- Pisoni DB. Perception of synthetic speech: The human listener as a cognitive interface. *Speech Technology*. 1982; 1:10–23.
- Pollack I, Pickett JM. Intelligibility of excerpts from conversational speech. *Language and Speech*. 1963; 6:165–171.
- Salasoo, A.; Pisoni, DB. Research on Speech Perception, Progress Report No 7. Bloomington, IN: Speech Research Laboratory, Indiana University; 1981. Perception of open and closed class words in fluent speech; p. 187-195.
- Stanovich KE, West RF. The effect of sentence context on ongoing word recognition: Tests of a two-process theory. *Journal of Experimental Psychology: Human Perception and Performance*. 1981; 7:658–672.
- Swinney D. Lexical access during sentence comprehension: (Re)consideration of context effects. *Journal of Verbal Learning and Verbal Behavior*. 1979; 18:645–659.
- Swinney, DA. The structure and time-course of information interaction during speech comprehension: Lexical segmentation, access, and interpretation. In: Mehler, J.; Walker, ECT.; Garrett, M., editors. *Perspectives on mental representation: Experimental and theoretical studies of cognitive processes and capacities*. Hillsdale, N.J: Erlbaum; 1982.
- Tyler LK, Marslen-Wilson WD. The on-line effects of semantic context on syntactic processing. *Journal of Verbal Learning and Verbal Behavior*. 1977; 16:683–692.
- Tyler LK, Marslen-Wilson WD. Conjectures and refutations: A reply to Norris. *Cognition*. 1982a; 11:103–107. [PubMed: 7198950]
- Tyler, LK.; Marslen-Wilson, WD. Speech comprehension processes. In: Mehler, J.; Walker, ECT.; Garrett, M., editors. *Perspectives on mental representation: Experimental and theoretical studies of cognitive processes and capacities*. Hillsdale, N.J: Erlbaum; 1982b.

Fig. 1.

Sample speech spectrograms of forward-gated and backward-gated sentences. Increasing signal duration is shown left-to-right (forward gating) and right-to-left (backward gating), respectively, for the target words of one test sentence.

Fig. 2. Identification points for words in meaningful and semantically anomalous sentences expressed as milliseconds of signal duration in each sentence position in Experiment 1. Forward-gated and backward-gated words (triangles and squares) are shown for each sentence context and in isolation (filled and open symbols). The measured physical durations of each target words at each sentence position are marked with X's and dotted lines.

Fig. 3.
Identification points for words in meaningful and semantically anomalous sentences expressed as proportions of the actual measured word duration for each sentence position in Experiment 1.

Fig. 4. Number of incorrect word responses generated for forward-gated and backward-gated words (triangles and squares) in each sentence position for the meaningful and semantically anomalous sentences, based on data from 42 and 45 subjects, respectively, in Experiment 1.

Fig. 5. Probability of correct identification as a function of signal duration for words in meaningful and anomalous sentences in Experiment 2 in the top and bottom panels, respectively. Triangles and squares represent forward-gated and backward-gated conditions, respectively. The curves cross the broken lines (.50 probability of correct identification) at the identification thresholds.

Fig. 6. Comparisons of identification performance as a function of signal duration for words in the meaningful and anomalous sentences (top and bottom panels, respectively). Data from successive presentations in Experiment 1 are shown as solid lines and data from single presentations in Experiment 2 are shown as broken lines.

Fig. 7.

Growth functions showing the distributions of word responses for increasing amounts of signal duration in the meaningful and anomalous contexts in Experiment 2. Data are expressed as proportions of the total responses for each gate duration. X's, triangles, and squares represent correct identification responses, incorrect responses based on acoustic-phonetic information, and syntactic sources, respectively; forward-gated and backward-gated conditions are shown in the left- and right-hand panels, respectively.

TABLE 1

Incorrect Response Distributions in Experiment 1 as Percentages of Total Errors

Response	<u>Meaningful context</u>		<u>Isolated control</u>	
	Forward	Backward	Forward	Backward
Blank	83.2	82.2	88.6	87.0
Acoustic – phonetic	8.5	8.5	7.9	9.4
Syntactic	4.9	7.1	2.0	2.0
Other	3.4	1.3	1.5	1.6

Response	<u>Anomalous context</u>		<u>Isolated control</u>	
	Forward	Backward	Foward	Backward
Blank	85.4	86.9	87.5	88.5
Acoustic – phonetic	10.8	8.7	9.6	9.1
Syntactic	2.7	3.2	1.2	1.1
Other	1.1	1.2	1.7	1.3

TABLE 2

Incorrect Response Distributions in Experiment 2 as Percentages of Total Errors

Response	<u>Meaningful context</u>		<u>Anomalous context</u>	
	Forward	Backward	Forward	Backward
Blank	80.3	74.4	60.3	72.7
Acoustic-phonetic	12.8	14.1	31.0	20.8
Syntactic	4.4	8.9	5.0	4.2
Other	2.5	2.6	3.7	2.3