EVIDENCE FOR EARLY EFFECTS OF SENTENCE CONTEXT ON WORD SEGMENTATION

Saskia te Riele and Hugo Quené

Utrecht Institute of Linguistics OTS, Utrecht University

ABSTRACT

The present paper focuses on the segmentation of two-word phrases containing two closely competing lexical hypotheses. It is hypothesized that the bottom-up information, which also includes a mechanism called the Possible-Word Constraint, is explored first in segmenting these phrases. Non-sensory sentential information influences this process at a later stage and only shows an effect if the bottom-up information does not lead to one dominating interpretation. The results of the present experiment show that listeners can and do make use of contextual information at a relatively early moment, at which the two possible segmentations are both still active and the bottom-up information has not yet suppressed the acoustically inconsistent interpretation. Hence, it was concluded that both sensory and non-sensory information is employed to affect activation levels of competing lexical hypotheses at an early moment.

1. INTRODUCTION

In order to understand a connected speech utterance, a listener first needs to identify the separate words which it contains. Identifying the boundaries between the words is not a trivial task. Although acoustic boundary markers may be present in the speech input, these cues are not always reliable (Nakatani and Dukes, 1977). However, word segmentation may be facilitated by characteristics of the word recognition process itself. The speech input has been shown to activate multiple lexical hypotheses (Zwitserlood, 1989), which compete with each other for word recognition (McQueen, Norris & Cutler, 1994). This competition process results in one best matching segmentation which dominates all others (Norris, 1994, Elman & McClelland, 1986).

The present paper focuses on the segmentation of two-word phrases containing two closely competing lexical hypotheses. An example is the two-word phrase /zeˈfɛɪn/. When the pivotal consonant in this fragment is short and contains a single /f/, it can be segmented into *zee fijn* ('sea fine') or *zeef fijn* ('sieve fine'), i.e. into a phrase with a single or a double underlying boundary consonant. This ambiguity is a consequence of the phonological process of degemination, which causes one of two adjacent and identical consonants in the underlying representation to be deleted in the surface form. This ambiguity may disappear if the pivotal consonant is long, so that it will be perceived as a double consonant (Pickett & Decker, 1960). The input string then contains two boundary consonants and matches the segmentation *zeef fijn* best.

Hence, if a phrase contains a long pivotal consonant, the bottom-up support favors a segmentation with a double boundary consonant. However, this segmentation will receive much competition from the interpretation with a single boundary consonant. A mechanism that can aid inhibition of the incorrect segmentation *zee fijn* is the Possible-Word Constraint (PWC), which was recently implemented in the Shortlist model (Norris, McQueen, Cutler and Butterfield, 1997). This constraint first derives potential word boundaries from phonotactic constraints, strong syllable onsets and silences. Then, it penalizes lexical hypotheses that leave non-vocalic material dangling between two potential word boundaries. Hence, if we take the phonemic input string /zeˈffɛɪn/, the word candidate *zee* will be penalized since the segmentation *zee#f#fijn* leaves an unexplained segment in the input.

Beside this bottom-up information, segmentation may also be facilitated by non-sensory information, like the preceding sentence context. In the above mentioned example, both *zee* and *zeef* remain active for some time, and at this point sentence context information may start playing a role (Norris, 1986, Gaskell & Marslen-Wilson, 1997). The activation of the contextually appropriate lexical hypotheses will then be boosted or its recognition threshold will be lowered.

The present study investigates the roles of the bottom-up information, including the Possible-Word Constraint, and the top-down sentential information during word segmentation. Although most recent models of word recognition ascribe a role of contextual information to word recognition processes, it is also generally assumed that the bottom-up information will be explored first, leaving a late role for sentence context (Connine, 1987). Effects of sentential information only become apparent in cases where sensory information does not lead to a dominating segmentation. This assumption leads to the prediction that in phrases with a double boundary consonant, the listener will delay word recognition until the bottom-up information is analyzed and a final decision can be made. Sentence context does not influence the competition process before the bottom-up information comes in. However, in phrases with a short, single boundary consonant, in which no disambiguating bottom-up information is available, sentential context is used to arrive at one dominating interpretation.

In order to test this hypothesis, the two-word phrases with a short and a long boundary consonant are embedded in three types of sentence context: 1) a semantically neutral context, in which both interpretations are possible; 2) a biasing sentential context in which information in the entire sentence renders one interpretation more appropriate; and 3) a sentence containing a lexical prime associated with one interpretation of the two-word phrases (e.g. *strand* 'beach' in a sentence containing the fragment /zeˈfɛɪn/). Sentences always biased towards an open first word (i.e. *zee*). Hence, the experiment contains situations

with conflicting sensory and non-sensory information. In this way it will be possible to determine which source of information was the most important factor at the time of the decision.

Since the two possible segmentations differ in their first words, a task is required that measures which word is recognized, *zee* or *zeef*. Rhyme-monitoring, in which a subject presses a button on hearing a word in the sentence that rhymes with a previously specified cue word, is such a task (te Riele, Nooteboom & Quené, 1996). In the present experiment, cue words always rhyme with the contextually appropriate segmentation. Therefore, a response indicates recognition of a phrase with a single boundary consonant (e.g. *zee*). No response means that listeners did not recognize a phrase with a single consonant, and it is assumed that in these cases a phrase with a double consonant was recognized (e.g. *zeef*). A further advantage of rhyme-monitoring is that it allows for on-line measurements, reflecting processing time of the two-word phrases.

Given our assumptions that a long consonant is perceived as a double consonant, and that bottom-up information has priority over top-down information, we expect no responses when a stimulus fragment with a long boundary consonant is presented, neither in neutral nor in biasing contexts. Because a short consonant can be perceived both as a single and as a double consonant due to regular degemination, about 50% responses are expected if a phrase with a single boundary consonant is embedded in a neutral context. In the biasing contexts, a shift towards more contextually appropriate responses is expected.

2. METHOD

2.1. Materials

The stimulus material consisted of 9 potentially ambiguous two-word phrases of the type /zeˈfɛɪn/, which may be interpreted as either *zee fijn* or *zeef fijn* ('sea fine' or 'sieve fine'). The duration of the pivotal consonant in these phrases was manipulated and could be short (mean = 87 ms), long (mean = 188 ms), or in between (mean = 132 ms). The consonant durations were determined in a classification experiment in which the duration of the pivotal consonant in each phrase was varied from 50 to 210 ms. Subjects were asked to indicate whether they had heard a phrase with a single or a double consonant. The duration at which a single or a double consonant was perceived in at least 90% of the instances was used as the short or long boundary consonant duration, respectively. The duration at which a single boundary consonant was perceived in 50% of the cases was used as the intermediate duration. Stimulus phrases always consisted of monosyllabic and mono-morphemic noun-adjective combinations with a voiceless fricative /s/ or /f/ as the boundary consonant. All 3 versions of each stimulus phrase were embedded at the end of the three types of sentence context described in the introduction.

The filler material consisted of 45 sentences, containing target words which were also part of potentially ambiguous two-word phrases with manipulated boundary consonant durations. Filler phrases differed in structure from the stimulus phrases and occurred at various positions in the sentences. In 25% of the cases the cue word did not rhyme with the target word at all, for instance since the vowel in the cue and target word differed.

All sentences were read by a male native speaker of Dutch at a relatively fast speaking rate and recorded on DAT tape in a professional studio. The material was re-sampled at 22.05 kHz and stored in a computer. One realisation of each minimal pair was then sliced out and its boundary consonant duration was manipulated. The three created versions of each phrase were then embedded in the sentence contexts.

2.2. Design

Independent variables were crossed in a 3 (consonant durations) x 3 (types of sentence context) matrix, resulting in 9 conditions. Nine different experimental tapes were created, so that all stimulus phrases were presented only once to each subject. The 9 conditions were counterbalanced across the 9 tapes.

In order to gain information about which interpretation (*zee fijn* or *zeef fijn*) was recognized in each condition, both a cue word rhyming with *zee* (e.g. *fee* 'fairy') and with *zeef* (e.g. *scheef* 'crooked') should be used. However, this would double the number of conditions, and consequently the number of subjects. To reduce these numbers, only cue words rhyming with the contextually appropriate, open target words are used (e.g. *fee*).

2.3. Subjects and Procedure

180 Students of Utrecht University (20 in each tape) were tested individually in a sound treated booth; all were native speakers of Dutch with no reported hearing impairment. They were asked to press a button as quickly as possible on hearing a word in the sentence that rhymed with the cue word presented in advance. The experiment was controlled by a computer, and cue words were presented both visually and auditory and were followed by the sentence after 500 ms. Subjects could respond during a 3 second period. Responses made outside that period were registered as misses. The stimuli were presented over closed headphones at a comfortable listening level. Prior to the experiment, a practice session with 16 trials was presented.

3. RESULTS

Percentages hits were calculated for each subject and each item, where each hit indicates recognition of a word with an open first syllable (i.e. *zee*). These percentages were subjected to two analyses of variance, one over the percentages per subject (F1) and one over the percentages per item (F2). Factors in the analyses were consonant duration and type of context. Reaction times were measured from the onset of the first vowel in the two-word phrases. A subject and item analysis of variance were performed on the reaction times as well, again with consonant duration and type of context as the other factors.

Figure 1 presents the percentages of hits for each consonant duration in each type of context. The effect of consonant duration was significant ($F1(2,171) = 44.58$, $p < 0.001$, $F2(2,8) = 12.65$, $p < 0.001$). Hence, the bottom-up information and the

PWC influenced segmentation at the decision stage, so that the stimuli with a long consonant are less often interpreted as a phrase with a single boundary consonant than the stimulus fragments with a short consonant. However, a long consonant always led to a non-zero number of hits, especially in context, although this is not expected if the bottom-up information has priority.

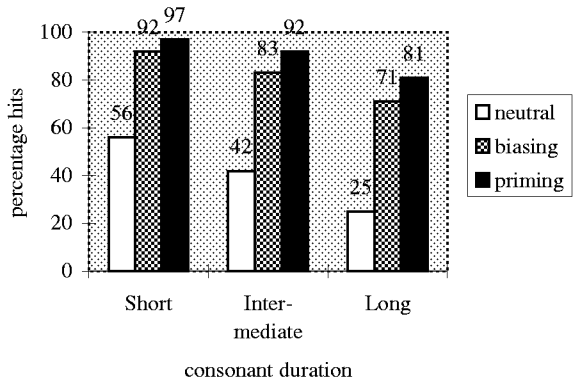

Figure 1: Percentages responses for each consonant duration in each type of context. Percentages in each bar are based on 180 responses (9 items * 20 subjects).

The figure also shows an effect of sentence context, which was significant in the analyses of variance ($F1(2,171) = 188.42$, $p < 0.001$, $F2(2,8) = 62.74$, $p < 0.001$). When the sentence context biases towards or primes the interpretation with the single boundary consonant, subjects indicated to have recognized this contextually appropriate segmentation more often as compared to the neutral context, which is also shown in a Tukey HSD test (neutral vs. biasing, $p < 0.001$, neutral vs. priming, $p < 0.001$). This shift towards more contextually consistent responses was expected in the short boundary consonant condition, since context is the only source of information that may help segmentation in these stimuli.

In the phrases with a long boundary consonant, where acoustic information and the PWC can be used first to arrive at a dominating segmentation, this shift was unexpected. It suggests that even if the bottom-up information is sufficient for segmentation, contextual information plays an important role in the decision process. Still, the interaction between type of context and consonant duration was not significant ($F1(4,171) < 1$, n.s. $F2(4, 8) < 1$, n.s.), which means that the effect of consonant duration did not become less important in context. Hence, contextual information did not completely suppress the effects of the acoustic information.

When considering the reaction times that go with the responses, a first thing to note is their great variance. Reaction times ranged from 47 to 3152 ms, where most responses are shorter than 500 ms. In order to use the acoustic information and the PWC, listeners need to have processed the entire boundary consonant, and probably part of the second word as well. But if the mean duration of the first vowel and boundary consonant are subtracted from these fast reaction times (223 ms in stimuli with a short consonant and 324 ms in stimuli with a long consonant), it appears that many responses were initiated at an early moment, at which the two possible lexical hypotheses are still active and the bottom-up information could not yet have suppressed the interpretation with the single boundary consonant.

Table 1 shows the mean reaction times for each experimental condition. In the neutral context, the reaction times are especially short in the phrases with a long boundary consonant, where the recognized word *zee* is inconsistent with the bottom-up information. These fast responses may be a consequence of a phonological priming strategy, which is sometimes used in the rhyme-monitoring task (te Riele, Nooteboom & Quené, 1996). Subjects then concentrate primarily on detecting words rhyming with the cue word. On hearing such a word, they will respond immediately, without waiting for durational information provided by the boundary consonant.

	Consonant duration			
	Short	Intermediate	Long	*Average*
Neutral	634	593	530	598
Biasing	539	541	520	534
Priming	508	471	481	487
Average	548	522	504	527

Table 1: Reaction times for each consonant duration in each type of context, measured from the onset of the first vowel in the phrases. Reaction times are based on the numbers of responses as shown in figure 1.

Reaction times are even shorter when the phrases are embedded in a biasing or priming context ($F1(2,170) = 11.15$, $p < 0.001$, $F2(2,16) = 4.75$, $p < 0.05$). This result suggests that the activation level of the contextually appropriate word was boosted early during processing, so that our subjects were led to give fast, contextually appropriate responses. This effect is further enhanced by phonological priming of the cue words. Context and cue word priming, thus, cause early decisions, that are made before the bottom-up information is fully explored and before it has led two one dominating segmentation.

4. DISCUSSION

In the introduction we hypothesized that listeners in segmenting potentially ambiguous two-word phrases give priority to the available bottom-up information, and try to make a decision on the basis of that type of information. Sentence context was assumed to be explored at a later moment, namely only after the bottom-up information suppressed one of the two possible segmentations (Connine, 1987). Only if the phrases are acoustically ambiguous, effects of sentence context would become apparent.

This hypothesis was tested by using two-word phrases with short and long boundary consonants. A long consonant indicates the presence of a double consonant in the input. As a consequence, a phrase with two boundary consonants can be segmented correctly on the basis of the acoustic information and the PWC alone. Hence, according to the above mentioned hypothesis, listeners will always recognize a two-word phrase

with an underlying double boundary consonant on hearing these phrases.

The results in the present experiment show an effect of the bottom-up information on the outcome of the segmentation and recognition process. On hearing fragments with a long boundary consonant, our listeners more often recognized a phrase with a double than a single boundary consonant. However, sentence context also influences recognition, and seems to affect activation levels at a moment when both possible segmentations are still active, and when the bottom-up information is not yet fully analyzed and available to suppress the incorrect segmentation.

The strong context effect that was obtained in the present experiment may be a consequence of the type of materials used. The disambiguating acoustic information in the two-word phrases arrives relatively late, namely only at the beginning of the second word. Also, the bottom-up PWC can only be employed after the second possible word boundary is located. As a consequence, segmentation must be delayed and both interpretations remain active for a relatively large period. During that period, sentence context has ample time to influence activation levels of the lexical hypotheses, which leads to a high number of fast, and contextually consistent but acoustically inconsistent responses.

This shift towards more contextually appropriate responses may have been expected in the contexts containing an associative prime. In these cases, the activation levels of associated words will be influenced directly through intra-lexical spreading of activation (Seidenberg, Waters, Sanders & Langer, 1984). The same kind of mechanism may be responsible for the effect of phonological priming of the cue word, where activation levels of rhyming (contextually appropriate) words may be boosted. However, the shift is also observed when the information in the entire sentence renders one interpretation more appropriate, and the context effect can not be explained by intra-lexical processes.

Hence, our findings indicate that an effect of higher-level sentential information during word recognition is not delayed until the bottom-up information is analyzed and one segmentation comes to be dominant. When two possible segmentations enter into competition, both the bottom-up information, including a mechanism like the Possible-Word Constraint, and the non-sensory sentential information influence activation levels of the competing lexical hypotheses.

6. REFERENCES

1. Connine, C.M. "Constraints on interactive processes in auditory word recognition: The role of sentence context", *Journal of Memory and Language, Vol. 26, 527-538, 1987.*

2. Gaskell, M.G. and Marslen-Wilson, W.D. "Integrating form and meaning: a distributed model of speech perception", *Language and Cognitive Processes, Vol. 12, 613-656, 1997.*

3. McClelland, J.L. and Elman, J.L. "The TRACE model of speech perception', *Cognitive Psychology, Vol. 18, 1-86, 1986.*

4. McQueen, J.M., Norris, D. and Cutler, A. "Competition in spoken word recognition: spotting words in other words', *J. of Exp. Psychology: Learning, Memory and Cognition, Vol. 20, 621-638, 1994.*

5. Nakatani, L.H. and Dukes, K.D. 'Locus of segmental cues for word juncture", *J. of the Acoustical Society of America, Vol. 62, 714-719, 1977.*

6. Norris, D. "Word recognition: context effects without priming", *Cognition, Vol. 22, 93-136, 1986.*

7. Norris, D. "Shortlist: a connectionist model of continuous speech recognition", *Cognition, Vol. 5, 189-234, 1994.*

8. Norris, D., McQueen, J.M., Cutler, A., and Butterfield, S. "The possible-word constraint in the segmentation of speech", *Cognitive Psychology, Vol. 34, 191-243, 1997.*

9. Pickett, J.M. and Decker, L.R. "Time factors in the perception of a double consonant", *Language and Speech, Vol. 3, 11-17, 1960.*

10. Riele, S. te, Nooteboom, S.G. and Quené, H. "Strategies used in rhyme-monitoring" In: H.T. Bunnel and W. Idsardi (eds.) *Proceedings ICSLP 96, Vol 1 , 90-94, 1996.*

11. Seidenberg, M.S., Waters, G.S., Sanders, M. & Langer, P. " Pre- and postlexical loci of contextual effects on word recognition", *Memory and Cognition, Vol.12, 315-328, 1984.*

12. Zwitserlood, P. "The locus of the effects of sentential-semantic context in spoken-word processing", *Cognition, 32, 25-64, 1989.*