Reading aloud a connected text: how the organization of topics affects sentence-final lengthening in English, French and Japanese

Caroline L. Smith, Lisa A. Hogan and Mami O. McCraw
University of New Mexico, Albuquerque, New Mexico, USA
E-mail: caroline@unm.edu

ABSTRACT

Languages differ as to which factors have the greatest effect on durational patterning, although some word- and sentence-level factors are known to be important in many languages. This experiment focuses on a factor operating over larger size units, and tests the extent to which the topic organization of a text affects durational patterns when speakers of English, French and Japanese read aloud. The pattern examined is sentence-final lengthening, which was measured by comparing the duration of words in sentence-final position in continuous text with their duration in sentence-medial position in a control sentence constructed for comparison. The results from all three languages provide evidence that topic structure affects the amount of final lengthening, but the languages appear to differ as to which types of transitions between topics favor more or less lengthening. Other results may be explained by structural differences among the languages.

1. BACKGROUND

Recent work in discourse prosody has documented multiple acoustic dimensions in which the acoustic signal reflects the organization of the material being spoken. The project reported on here is concerned with understanding the behavior of speakers reading aloud as an aid to improving synthesis. Earlier results on prosodic effects relative to discourse organization are summarized in [6]. More recent research has found differences in English in F0 and RMS amplitude, as well as durations, between words in discourse-final and non-final position [5], [7].

Previous research has focused on the role of intonation, particularly manipulation of pitch range in structuring discourse [1], [12]. Here we concentrate on the contribution of durational factors, and because our primary interest is in production, we look at how speakers' durations reflect the organization of content in a text being read aloud. Durational patterning on the sentence level is known to differ across languages (e.g., [2]), but less is known about cross-linguistic differences in durations over larger spans of speech (although see [3]). Our experiment facilitates comparison among the three languages studied by using the same methodology to construct and analyze similar materials read by native speakers of English, French and Japanese. To ensure that the texts were representative of the different languages, we chose separate texts published in each language.

All of the studies cited above report a relation between phonetic properties and the organization of the spoken content, but the methods used to specify discourse organization vary. One strategy is to compare sentences at different *positions* in a discourse, since the location of a sentence relative to the rest of its discourse context helps to determine its purpose vis-à-vis that context. Other studies, such as [3], [4], [10], have taken explicit theoretical approaches to identifying aspects of discourse organization which correlate with prosodic differences.

Theoretical models of discourse were less relevant in our study. Speech synthesis of appropriate discourse-contextual patterns requires only the identification of where those contexts are, not of the overall structure of the discourse. In addition, computers are not yet capable of generating sophisticated discourse analyses, so any methodology which uses such analyses cannot be incorporated into a synthesis system. Therefore we chose to adapt for written discourse a method which Nakajima and Allen [9] used to describe the relation between sequential utterances in a spoken discourse. In this approach, the transition from one sentence to the next is labeled as a Topic Continuation if the following sentence continued the topic, advancing the narrative; an Elaboration if the following sentence provided additional detail; or a Shift if the following sentence introduced new material. We created one additional category, Text Marker, analogous to discourse markers in speech. This label indicated that the following "sentence" overtly marked textual organization. In our texts, most Markers were numbers in sequences of instructions. These labels were assigned to sentences with reference to the nature of the transition from that sentence to the next, because it seemed more likely that the final word in the sentence would reflect the nature of the upcoming transition, rather than the topical transition from the preceding sentence.

2. METHOD

The goal was to determine how the nature of a topic boundary between sentences affects final lengthening, and to compare this effect across three diverse languages.

2.1. Text materials

In order to determine the amount of lengthening in a particular production of a segment or word, and control for segmental and word-level factors, we compared the

ISBN 1-876346-48-5 © 2003 UAB

duration of the same 'target' words as they occurred sentence-finally in a text and as they were produced in a controlled sentence-medial context. For each target word, a control sentence was constructed with the target word in sentence-medial position. All of the control sentences were of similar length, with the same number of syllables (or moras, in Japanese) before and after the target word for all control sentences in a given language. In English, the control sentences were designed to favor the production of the target word with or without a pitch accent depending on whether the presence or absence of accent seemed likely for that word where it occurred in the text.

Text selections were made on the basis of several criteria. Each text extract had to form a reasonably cohesive passage, and there had to be a variety of words in sentence-final position, to permit the measurement of lengthening across different segment and syllable types. The second criterion posed significant challenges in Japanese, a verb-final language which does not inflect for person or number, as most sentences within any written text tend to end with one of two endings, either the present or past tense verb suffixes. The English and French texts, and one of the Japanese texts, were very similar in style, being extracts from computer manuals instructing the reader in the use of some piece of software or equipment. The second Japanese text (referred to here as the "traffic" text) was a government document about road safety. It was chosen because it included a much wider range of words in sentence-final position than the other Japanese text. But the overall style of the traffic text was different from any of the other texts, and it included a number of Chinese characters which were initially unfamiliar to the readers. The Japanese experimenter discussed these with readers prior to recording, but their reaction to this text was clearly different from the other text (referred to here as the "internet" text). The English and French readers reported no difficulties with their texts.

2.2. Speakers and recording technique

Five native speakers of American English, three French speakers from France and three Japanese speakers were recorded reading these materials. Because the overall purpose of this project is to develop a model of naturally-occurring durational patterns for speech synthesis, the focus is on developing a detailed picture of the behavior of individuals, rather than averages over many speakers. Data from one speaker of each language are reported here: a male speaker of English, a male speaker of French and a female speaker of Japanese.

The recordings were made on a Sony Professional Walkman, using a Shure head-mounted microphone. Ten recordings were made of each speaker, with a mean interval between recording sessions of ten days for the English and French speakers, and seven days for the Japanese speaker. At each recording session, speakers were presented with a different order of the four sections (two texts and two sets of control sentences) that alternated text and control sentences, with the order of the control sentences randomized across sessions. Results from one text each for English and French and from both Japanese texts are reported here. The English text contained 60 sentences, the French text contained 78 sentences, and the Japanese texts contained 53 and 78 sentences respectively.

2.3. Analysis

The recordings were digitized on a Kay Elemetrics CSL system; all measurements were made with CSL. The acoustic duration of the target words was measured using the speech waveform and where necessary a spectrogram to identify acoustic landmarks. In addition to the duration of the entire target words, additional measurements of smaller word-final units were also made: the final rime (vowel plus following consonants) in English, the final syllable in French, and the final mora in Japanese. These were measured in order to see how much of the sentence-final lengthening was localized at the very end of the word. The syllable and mora are generally considered to be the basic timing units of French and Japanese, and previous research (e.g. [14]) has shown that final lengthening in English is concentrated in the final rime.

Numerical results were processed in Excel, then statistical analyses were run on the data in each language separately. Descriptive statistics were performed using StatView; ANOVAs were calculated using the Mixed procedure in SAS, which is appropriate for analyses involving both fixed factors and repeated measures. The significance level used was .05. The most important factor, tested in all three languages, was Topic Transition Type. For English, the ANOVA also included fixed factors of word accent (with or without pitch accent) and syllable stress (primary stress or reduced); for Japanese, where two texts were used, text identity was also a fixed factor. Post-hoc Scheffé's tests were used to identify which levels of Topic Transition Type differed significantly.

2.3.1. Calculation of amount of lengthening

The mean durations of the words in control sentences were calculated using a technique similar to that used by Wightman et al. [14] and other studies. The duration of each target word in a text sentence was pseudo-normalized to permit comparison among the different words. If d-t is the duration of one reading of a target word in the text, and μ-s is the mean duration of the same target word across the ten readings of the control sentences, then the pseudo-normalized duration was

$$\text{d-t/norm} = (\text{d-t} - \mu\text{-s}) / \sigma\text{-s} \qquad (1)$$

In English, the mean and standard deviation used were those for the accented or unaccented productions from the control sentences, depending on whether the reading whose duration was being 'normalized' was accented or unaccented. In the rest of this paper, all references to duration refer to this pseudo-normalized duration. (Note that this is not the same as a z-score, because the mean and standard deviation being used are calculated over a sample distinct from the values being normalized.)

2.3.2. Topic transition types

The types of transitions between each sentence in the texts were labeled by native speakers of the appropriate languages, none of whom had participated as speakers. All labelers were graduate students or researchers in linguistics, and all were given the same set of instructions (in English). There were five labelers for English, four for French and three for Japanese working independently. In cases of disagreement, consensus labels were assigned by choosing the label preferred by the majority. In two of the Japanese sentences there was no majority preference and these were discarded; the French labelers disagreed more often and in half of the sentences a label was used which was selected by only two of the four labelers.

3. RESULTS

The number of readings of sentences classified in each type of topic transition is listed in Table 1. The values listed for Japanese in this table combine the totals from both texts. However, the distributions of topic types differed in the two texts; for example, all Japanese Text Markers occurred in the traffic text.

Topic Transition Type	English	French	Japanese
Topic Shift	69	70	175
Topic Continuation	251	309	434
Elaboration	149	190	221
Text Marker	57	120	38

Table 1: Count of tokens of each topic transition type.

In all three languages, the ANOVAs showed a significant effect of Topic Transition Type on lengthening of the sentence-final word and the smaller units, but the consequences of this effect varied considerably among the languages, so each language will be described separately.

3.1. English

The ANOVA showed significant main effects of Topic Transition Type and syllable stress, and interactions between these factors as well as between these factors and word accent. The effect of Topic Transition Type is graphed in Figure 1. Post-hoc tests showed a significant difference only between Elaborations, with the least amount of lengthening, and Topic Shifts, with the most.

Lengthening in the final rime also showed significant main effects of Topic Transition Type and syllable stress, and interactions between each of these factors and word accent. The amount of lengthening in the final rime was significantly greater with Topic Shifts than with any of the other Transition Types, and Continuations had more lengthening than Elaborations or Markers. A fuller account of the results for English can be found in [11].

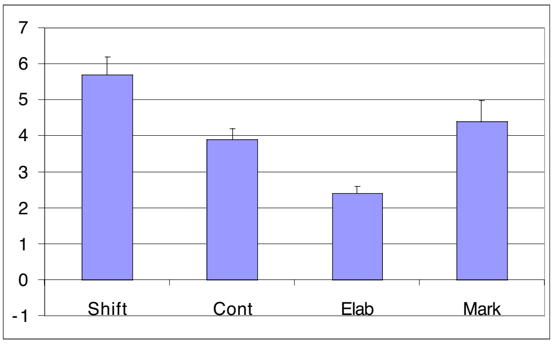

Figure 1. Amount of lengthening in the sentence-final words in English.

3.2. French

As in English, there was a significant main effect of Topic Transition Type in the ANOVA. All differences were significant. Topic Shifts had the greatest amount of lengthening. Similar results were found for lengthening of the final syllable alone.

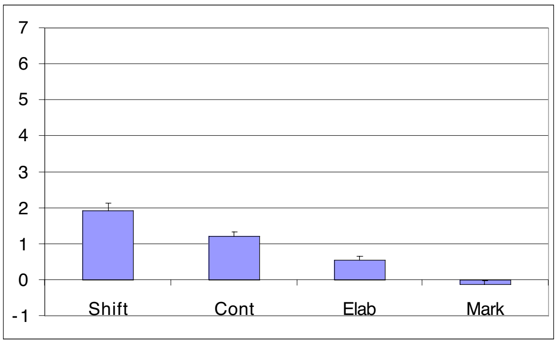

Figure 2. Amount of lengthening in the sentence-final words in French.

Results for the French speaker differed from those of the English speaker primarily in the magnitude of lengthening observed. As is apparent from comparison of Figures 1 and 2, there was much less lengthening in French, so little that in Topic Elaborations and Text Markers, the amount of lengthening was not significantly different from zero in a one-sample sign test. The most likely explanation is that in French, the sentence-medial words in control sentences, whose duration was the basis for calculation of lengthening, occurred in a prosodically prominent position at the end of an accent group. Words in this position tend to be lengthened [8]. Therefore, when the duration of these words was compared to their sentence-final duration, there was little increased duration sentence-finally, because in both cases the word was at the end of the accent group, which caused lengthening in both environments.

3.3. Japanese

The magnitude of lengthening in Japanese was comparable to what was found in English, as can be seen by comparing Figure 3 and Figure 1. In the ANOVA for Japanese, there were significant main effects of Topic Transition Type and of text identity and an interaction of Transition Type and text identity.

ISBN 1-876346-48-5 © 2003 UAB

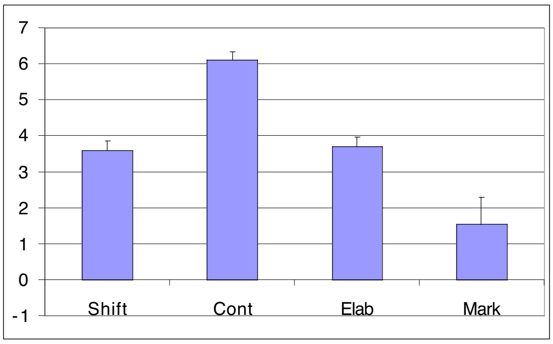

Figure 3. Amount of lengthening in the sentence-final words in Japanese.

Overall, post-hoc tests showed that Continuations were significantly longer than Elaborations or Shifts. Text Markers were the shortest, but this result should be treated with particular caution, as there were very few tokens of these. Because of the significant interaction with text identity, the Japanese texts were also analyzed separately. Both of them differed from English and French in that at least one other Transition Type had more lengthening than Topic Shifts. However, the two texts also differed substantially. In the traffic text, as in the overall results shown in Figure 3, Continuations had more lengthening than Elaborations, but the reverse was true in the internet text. It appears that the different styles of writing in the texts led to different reading styles on the part of our speaker. The analysis relies on the labelers' determination of the text's structure. If their interpretation did not match the speaker's, the labeling of Topic Transition Types may not coincide with how the speaker actually read the text.

These results suggest that Japanese differs from English and French with regard to Topic Shifts, but the differences between the two Japanese texts make it difficult to conclude exactly how the languages differ.

4. CONCLUSIONS

Our initial expectation was that English would show more effect of topic structure on timing than French and Japanese. English displays various kinds of temporal variation, such as vowel reduction due to stress, that French and Japanese do not. The effect of discourse structure on intonation may be more subtle in Japanese than in English [13], but our results for lengthening in Japanese were comparable to those for English.

When only one speaker per language has been studied, it is impossible to distinguish between speaker-specific behavior and differences that are truly a consequence of the speakers' different languages. Nonetheless, at least some of our results seem likely to relate to differences between the languages themselves: the best explanation for the small amount of lengthening in French relates to the prosodic structure of French, although there may also be idiosyncratic tendencies favoring more or less lengthening by individual speakers or in specific languages. All three of the languages clearly showed effects relating to topic structure, suggesting that this line of research can lead to better understanding of the relation between discourse organization and phonetic properties.

ACKNOWLEDGMENTS

This work was supported by NSF grant BCS-9983106.
We also thank our speakers and labelers for participating.

REFERENCES

[1] G. Ayers, "Discourse functions of pitch range in spontaneous and read speech," *OSU Working Papers in Linguistics*, vol. 44, pp. 1-49, 1994.

[2] G. Fant, A. Kruckenberg and L. Nord, "Durational correlates of stress in Swedish, French and English," *Journal of Phonetics*, vol. 19, pp. 351-365, 1991.

[3] Y-J. J. Fon, *A cross-linguistic study on syntactic and discourse boundaries in spontaneous speech*, Ph.D. dissertation, Ohio State University, 2002.

[4] B. Grosz and J. Hirschberg, "Some intonational characteristics of discourse structure," in *Proc. of the 2nd ICSLP*, pp. 429-432. Banff Alberta, 1992.

[5] R. Herman, "Phonetic markers of global discourse structures in English," *Journal of Phonetics*, vol. 28, pp. 466-493, 2000.

[6] J. Hirschberg, "Studies of intonation and discourse," in *Proc. ESCA Workshop on Prosody*, Working Papers, Dept. of Linguistics and Phonetics, Lund, Sweden, vol. 41, pp. 90-95, 1993.

[7] J. Hirschberg and C. Nakatani, "A prosodic analysis of discourse segments in direction-giving monologues," in *Proc. of the 34th Annual Meeting of the ACL*, pp. 286-293, 1996.

[8] A. Lacheret-Dujour and F. Beaugendre, *La prosodie du français*, Paris: CNRS Editions, 1999.

[9] S. Nakajima and J. Allen, "A study on prosody and discourse structure in cooperative dialogues," *Phonetica*, vol. 50, pp. 197-210, 1993.

[10] L. Noordman, I. Dassen, M. Swerts and J. Terken, "Prosodic markers of text structure," in *Discourse Studies in Cognitive Linguistics*, K. van Hoek, A. Kibrik and L. Noordman, Eds., pp. 133-148. Amsterdam: John Benjamins, 1999.

[11] C. Smith, "Topic transitions and durational prosody in reading aloud: production and modeling," submitted.

[12] J. Venditti, *Discourse structure and attentional salience effects on Japanese intonation*, Ph.D. dissertation, Ohio State University, 2000.

[13] J. Venditti and M. Swerts, "Intonational cues to discourse structure in Japanese," in *Proc. of ICSLP 96,* vol. 2, pp. 725-728. Philadelphia, 1996.

[14] C. Wightman, S. Shattuck-Hufnagel, M. Ostendorf and P. Price, "Segmental durations in the vicinity of prosodic phrase boundaries," *JASA*, vol. 92, pp. 1707-1717, 1992.