

Naïve listeners' perceptions of French prosody compared to the predictions of theoretical models

Caroline L. Smith, University of New Mexico

Most models of prosody have been developed from relatively small collections of recordings, analyzed by a small number of linguists. A useful counter-balance to this strategy is available in a methodology that employs a large number of untrained listeners to provide an impressionistic perceptual transcription of prosodic structure (Mo, Cole and Lee 2008; see also Streefkerk et al., 1997, Buhmann et al. 2002). This methodology gives insight into which prosodic distinctions are highly salient, rather than which may be detectable under careful examination of the acoustic signal. Mo et al.'s (2008) results for conversational American English demonstrated a good level of agreement among listeners, despite a fair amount of variation. The success of Mo et al.'s study suggests that this methodology is valid and could be exploited to investigate other languages. The listeners in Mo et al. (2008) performed two different tasks, marking prominent words and marking boundaries between groups of words. In English, the locations of prominences and boundaries are assumed to be unrelated; Mo et al. do not report any explicit comparison of the locations marked in the two tasks. In contrast to this lack of correlation in English, in French these two dimensions are closely connected. Primary phrasal prominence is described as "un accent de groupe" (DiCristo 1999:164), since it occurs at the end of a "groupe accentuel", the smallest phrasal unit (Pasdeloup 1990; DiCristo 1999). Jun and Fougeron (2002) say that "the final full syllable of a word is realized with longer duration and higher intensity [that is, as more prominent] only if it is the last full syllable of a phrase." (p. 147) In other words, prominence is a by-product of the presence of a phrasal boundary, not an inherent property of a word (Vaissière 2002). We might thus expect prominences to be identified on the word preceding a boundary.

Method

This hypothesis was tested in an experiment conducted along the same lines as Mo et al.'s study. Two types of speech materials were used. Ten extracts were prepared from recordings of a map task experiment. These extracts consist of spontaneous task-directed speech, taken from the portions of a conversation during which only one speaker was speaking; the speakers are ten female undergraduates at a Paris university, and the extracts are between 13 and 24 seconds in length. The second group of ten extracts were taken from a discussion broadcast on a current affairs program on the France Info radio station. These also consist of single-speaker passages of spontaneous conversational speech, but the speakers are journalists and public figures. Their conversation was recorded for broadcast and thus illustrates a more formal register. There are two extracts from each of five speakers. These extracts last from 26 to 53 seconds. Orthographic transcriptions of the extracts, without punctuation, were prepared by the experimenter (a fluent non-native speaker of French), then edited by a phonetically-trained native speaker. Three additional extracts (two from the map task conversations, one from a similar radio program) were also prepared to serve as practice samples. Forty-six French listeners without advanced training in phonetics or prosody were recruited, mostly undergraduate students in Linguistics. Different listeners participated in the experiment in different settings at universities in Paris and Lyon: some were tested in groups of 5-17 in a classroom, others individually or in groups of two or three in a sound-attenuated room. They were presented with a packet containing instructions and the printed transcriptions of the practice and test extracts. They marked all their responses on these print-outs. Each listener performed one of two tasks. 22 listeners were instructed to underline words that

were highlighted (“mis en relief”); 24 listeners were asked to mark a vertical line between words at locations where they perceived a boundary between two phrasal units (“syntagmes”, defined as groups of words that form a single unit). All listeners heard the extracts in the same order. They practiced their task first on two map task extracts, then responded to ten map task extracts, then did a practice with an extract from a radio interview, then responded to the ten extracts from the radio broadcast. With the exception of one listener in the boundary-marking group who failed to follow directions for the map task extracts, all responses to the experimental samples were retained, and coded in Excel spreadsheets.

Results: Inter-listener agreement

Agreement among listeners was assessed using a modified form of Cohen’s kappa (Brennan and Predinger 1981) suitable for tasks in which the raters are not constrained as to how many items they assign to each category (“free marginals”), and calculated using the Online Kappa Calculator (Randolph 2008). Kappa values were calculated for each extract, pooling across all the listeners in each of the two groups. For marking of prominence, kappa ranged from .52 to .79, with a mean of .69 across the 20 extracts. For boundary marking, kappa ranged from .75 to .88 with a mean of .83. Because of the different calculation method, kappa values reported here are not directly comparable to those in Mo et al. (2008) for English or Buhmann et al. (2002) for Dutch. Nonetheless, it is striking that in the present study, as in the earlier ones, higher rates of agreement were obtained among those participants who marked boundaries than among those marking prominence. This is especially surprising given the usual assumption that in French, prominence derives from the occurrence of a boundary, and suggests that listeners’ perceptions of the two aspects of prosody are not in fact derived from the same information. Also comparable to Mo et al.’s results was the tendency to mark boundaries less often than prominences. This trend was observed in the current experiment for eight of the ten map task extracts, and six of the broadcast extracts (using medians rather than means to reduce the influence of outliers). The global median was one boundary marked every 9.9 words, and one prominence every 8.6 words, with more frequent marking in the map task extracts.

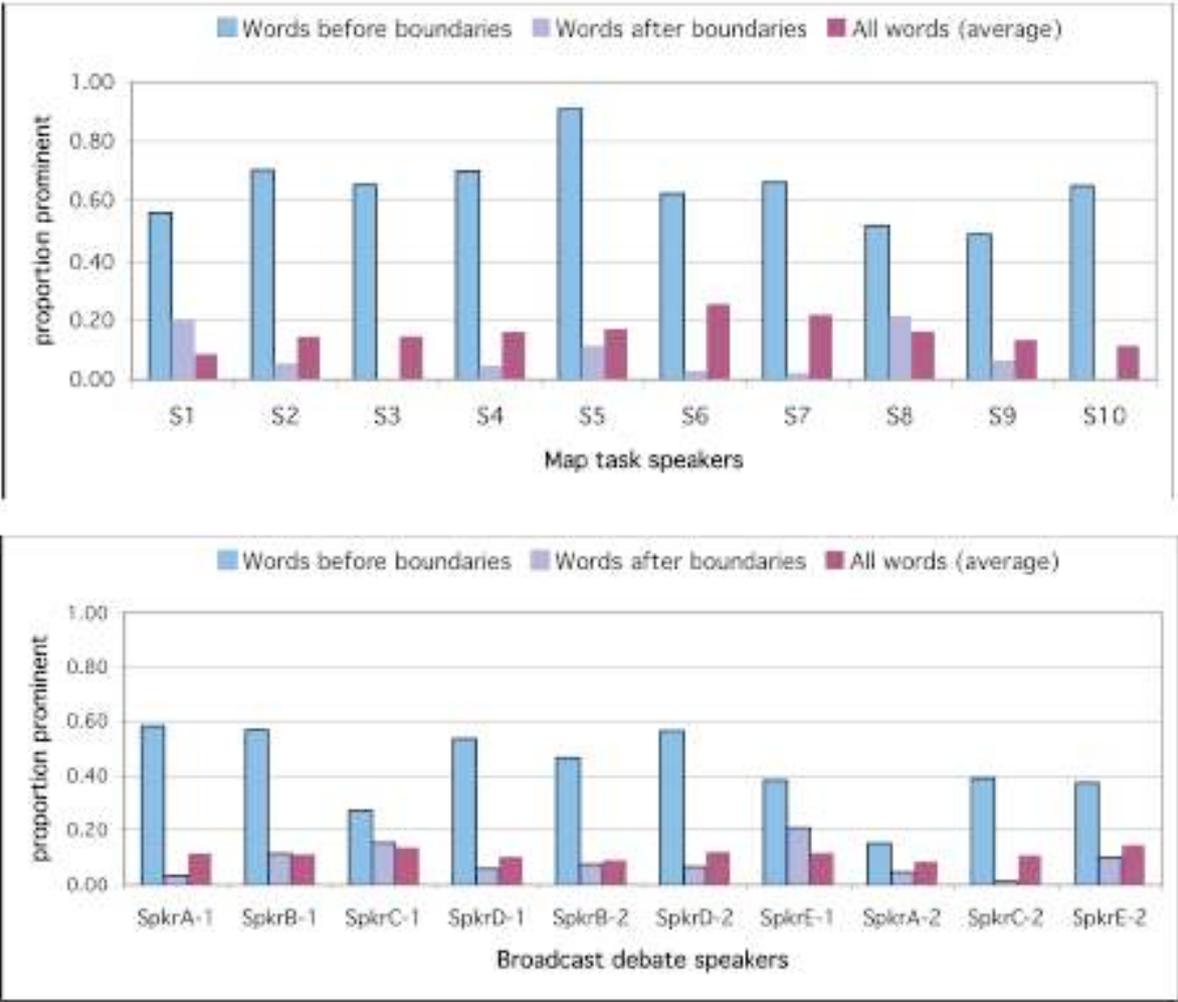
Results: Location of perceived prominences relative to perceived boundaries

In order to test the hypothesis of a correlation between the locations of prominences and boundaries, each word in each extract was assigned a prominence score and a boundary score equal to the proportion of listeners who marked it as having prominence, or as being followed by a boundary. Two analyses were then carried out using these scores. In the first of these, the correlation between the two sets of scores was calculated for each extract. The large number of words receiving no marking (and hence scores of 0) would tend to inflate the correlation between the two scores, so all words that received two scores of 0 were excluded from calculation of the correlation. The correlation values reported here are thus conservative estimates. Across the ten map task extracts, correlations averaged .67 with a standard deviation of .08. Across the ten broadcast extracts, correlations averaged .50 with a standard deviation of .17. The map task extracts thus appear to offer good support for the hypothesis, while the much more variable broadcast extracts do not consistently support it. The broadcast speakers were working hard to make their point, and thus may have exploited more marked prosodic options than those used by ordinary speakers. Nonetheless, they apparently were not emphasizing the phrase-initial word, which is a frequent location for emphasis in broadcast and political speech (Di Cristo 1999), since the correlations between boundaries and prominence on the following word were very low and usually negative (mean = -.23).

In the second analysis, all locations in the extracts were identified that received a boundary score of at least 0.67, that is, at least 67% of the participants had marked that location as a boundary.

This criterion was arbitrary but indicates a substantial consensus. The number of locations so identified ranged from two to five among the map task extracts, and from five to eleven in the broadcast debate extracts. The prominence scores of the words before and after these boundaries were then obtained in order to determine whether the words in these positions have prominence scores that diverge from the average for the extract. Figure 1 shows that, as expected, words before boundaries received much higher prominence scores than average. Words after boundaries, on the other hand, received much lower prominence scores, in fifteen cases lower than the average for all words.

Figure 1. For each extract, prominence scores for words before and after boundaries, and average prominence scores.



Discussion

The results provide strong support for models of French phrasal structure that claim that

phrasal groups end with a prominence. Of course, the “phrases” indicated by listeners in this study do not correspond to the phrases that would be predicted by any model; this discrepancy is immediately apparent because the listeners marked a boundary approximately once every ten words, whereas many models predict that each lexical word will generally be in a separate phrasal unit.

Most models assign prominence to the final full-vowel syllable of the phrase-final word. Because the listeners in this study were doing their annotating very rapidly, at the real-time speed of the speech, it was not feasible to ask those who marked prominence to indicate which specific syllable they were hearing as prominent. Thus the data here cannot precisely corroborate this claim of the models. It remains to be explained why so few phrase-initial words were marked as prominent, given that this is the predicted location of secondary prominence. One reason may be that in some of the phrases the initial word is a grammatical word and thus less likely to receive prominence. A more complete test would require consideration of the lexical/grammatical status of the words in different phrasal positions.

The greater agreement among listeners in boundary-marking than prominence suggests a greater similarity between the results of this study and earlier studies of English and Dutch than might have been expected. It could suggest that the perception of French listeners may resemble perception by listeners in these other languages more than some models suggest. Nonetheless, the correlation between the locations of prominence and boundaries is specific to French. These findings coincide well with Dahan (1996), who observed a similar role for rhythmic groups in perception by French and Dutch listeners, but with French listeners also exploiting French specific processes.

References

- Brennan, R. Prediger, D. (1981). Coefficient Kappa: Some uses, misuses and alternatives. *Educational and Psychological Measurement* 41, 687-699.
- Buhmann, J., Caspers, J., van Heuven, V., Hoekstra, H., Martens, J.-P., Swerts, M. (2002). Annotation of prominent words, prosodic boundaries and segmental lengthening by nonexpert transcribers in the Spoken Dutch Corpus. *Proceedings of LREC 2002* (Las Palmas), 779-785.
- Dahan, D. (1996). The role of rhythmic groups in the segmentation of continuous French speech. *Proceedings of the 4th ICSLP*, Philadelphia, Vol. 2, 1185-1188.
- Di Cristo, A. (1999). Vers une modélisation de l'accentuation en français : première partie. *Journal of French Language Studies* 9, 143-179.
- Jun, S.-A., Fougeron, C. (2002). Realizations of accentual phrase in French intonation. *Probus* 14, 147-172.
- Mo, Y., Cole, J., Lee, E. (2008). Naive listeners' prominence and boundary perception. *Proceedings of Speech Prosody 2008* (Campinas).
- Pasdeloup, V. (1990). *Modèles de règles rythmiques du français appliqué à la synthèse de la parole*. Thèse de doctorat, Université de Provence.
- Randolph, J.J. (2008) Online Kappa Calculator. Retrieved April 14, 2009, from <http://justus.randolph.name/kappa>
- Streefkerk, B., Pols, L., ten Bosch, L. (1997). Prominence in read aloud sentences as marked by listeners and classified automatically. *Proceedings of IFA* (Amsterdam), 101-116.
- Vaissière, J. (2002). Cross-linguistic prosodic transcription: French versus English. In N.B. Volskaya, N.D. Svetozarova & P.A. Skrelin (eds.), *Problems and methods in Experimental Phonetics*. In honour of the 70th anniversary of Prof. L.V. Bondarko. St. Petersburg: St. Petersburg State University. Pages 147-164.