Fundamental Frequency and Other Prosodic Cues to Topic Structure

Margaret Zellers, Research Centre for English and Applied Linguistics, Univ. of Cambridge
Brechtje Post, Research Centre for English and Applied Linguistics, Univ. of Cambridge

Studies of prosody have often neglected to examine structure above the level of the utterance or sentence, yet it is becoming more and more apparent that prosody is an essential cue to the organization of longer discourses. Just as prosody helps to hold discourse together within utterances by indicating information that is new or noteworthy in some way, it also joins utterances together into groups which share the same discourse topic, or are "about" the same thing, in discourse. However, the semantic content of this kind of joining function is still unclear in that some theories propose a gradient interpretation of topicality, while others take a more categorical approach. An example of a topic structure theory with a gradient view is the hierarchical model of Grosz & Sidner (1986). Since the simple hierarchy they propose is potentially infinitely recursive, this theory lends itself to a gradient pattern of variation, where variation depends upon the amount of active information. In this type of theory, the notion of a "topic" as an element of the discourse structure may not be relevant at all, since the topicality relationships would be based primarily upon the dominance and subordination relationships between utterances. An alternative is proposed by Nakajima & Allen (1993) and Wichmann (2000) who suggest that topic structure can be described by four discontinuous categories related to the relationships between information contained within individual utterances. This is not totally unlike a hierarchical approach in that utterances may dominate or be subordinate to one another; however, in this type of theory the semantic relationships between utterances become relevant in addition to the hierarchical structure. Both theories allow for a great degree of detail in the organization of discourse; that is, very fine differences in how information relates to the topic may be expressed. They also assume (to differing degrees) that topic is a delimited characteristic; that is, it has an identifiable beginning and/or end in discourse. This assumption may be difficult to reconcile with conversation analysts' finding for spontaneous speech that topic changes may be gradual and difficult to bound (cf. Jefferson 1984).

In view of these potentially mutually inconsistent theories, the search for prosodic correlates of topic structure is difficult to separate from the search for an appropriate theory to describe topic structure. Fortunately, by accounting separately for other known influences on prosody, it is possible to examine the variation that remains with an eye to the topic structure of the discourse at hand. Zellers (2009) used a read text which had been constructed to control for the segmental structure of items being compared directly. Utterances in the text conformed to topic structure categories similar to those proposed by Wichmann (2000); this allowed for the testing of both category-based and hierarchy-based theories on the same materials. The text was read aloud by eighteen native speakers of Standard Southern British English (SSBE) and the recordings were annotated for segmental and prosodic characteristics. On the basis of these recordings, Zellers (2009) found evidence that the span of fundamental frequency (F0) falls correlates with those topic structure categories to an extent, although it is unclear whether a category analysis or a hierarchical analysis is more appropriate to describe the data. Contrary to the findings of Wichmann (2000), in these data there was no correlation between F0 peak timing and topic structure. Zellers et al. (2009) therefore examined variation in the timing of F0 peaks in relation to topic structure in the same data, and found that the phonological model used to describe the intonation affected the availability of this cue as an indicator of topic structure. A model where all peaks were categorized as the same pitch accent accounted for most peak timing variation through correlation with segmental factors, while a model which divided the same data into two pitch accent categories used peak timing

to differentiate between these two categories, one of which was potentially used to signal more "significant" information and was preferred in new topics as a result. Use of the one-category model would privilege a binary description of topic structure (i.e., material is either a new topic beginning or not), while use of the twocategory model allows for the presence of more levels of structure, but can say little about whether these levels better correspond to a hierarchical or a gradient model, since the F0 peak timing variation would appear to be only secondarily linked with the topic structure.

Although these production data show that speakers consistently vary F0 in relation to topic structure, prosody consists of more than simply F0. The presence of structured variation in F0 suggests that it is potentially still a useful cue when used in conjunction with other prosodic (and probably lexical) signals. Furthermore, individual speakers vary in the way they use these kinds of F0 variation. In order to gain a better understanding of the interaction between various prosodic cues, it is useful to examine data for individual speakers separately. A case study on two representative speakers from the read-speech production studies above shows that a speaker who takes less advantage of the topic signaling provided by variations in F0 fall span still shows topic-structure related variation. Instead of making use of F0 modulation for all categories or levels, this speaker appears to be signaling topic structure by varying her speech rate so that new topic utterances begin at a faster rate (cf. Figure 1). By combining these two cues, three topic structure 'groups' can be distinguished that could not be distinguished by her F0 fall span variations alone. Conversely, a speaker who uses F0 spans as signals to distinguish three groups does not vary her speech rate to signal new topics as distinct from other topic structure categories (cf. Figure 2). Of eighteen speakers in the original production study, the productions of eleven pattern with speaker two, and the productions of five with speaker one. The remaining two speakers in the study do not clearly fall into one or the other of these groups, but as they are the speakers with the smallest number of available data points, this is perhaps unsurprising.

The results of these case studies show the importance of examining a variety of different cues as signals of topic structure. Whereas one cue on its own might suggest a particular interpretation of the topic organization of spoken discourse, when we observe the cues in concert as the speaker produced them, a different pattern may emerge. In order to gain an accurate picture of the relevant discourse structures, further studies must take into account as many possible cues as possible, as well as accounting for individual differences in their use.
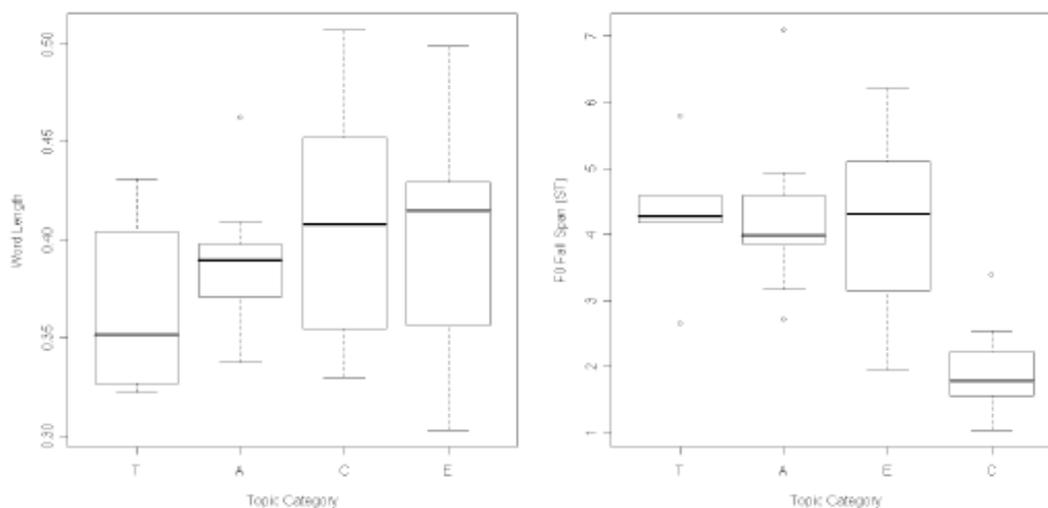

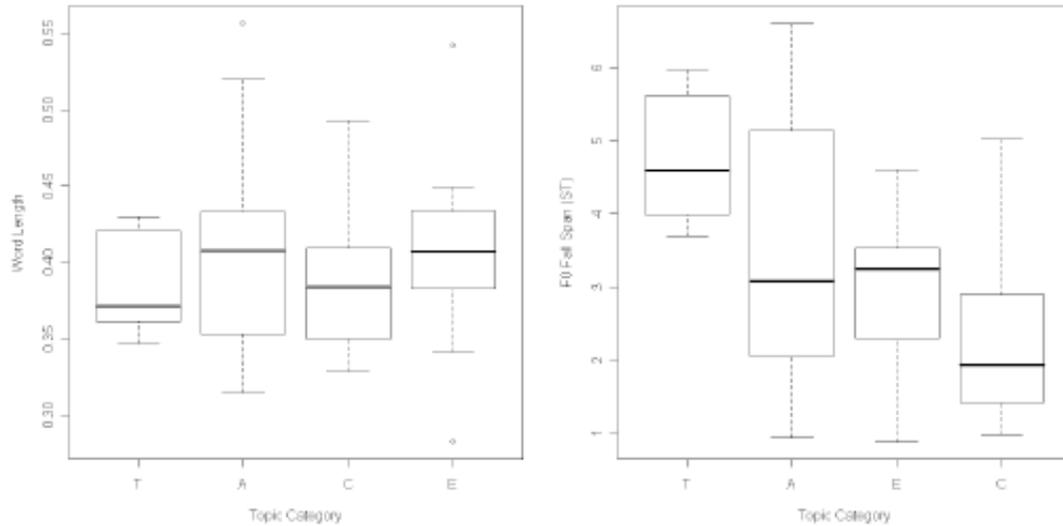Figure 1: Speech rate (left) and F0 fall span (right) for speaker 1.

Figure 2: Speech rate (left) and F0 fall span (right) for speaker 2

References

Grosz, B. and Sidner, C. (1986) "Attention, intentions, and the structure of discourse", *Computational Linguistics* 12(3):175-204.

Jefferson, G. (1984) "On stepwise transition from talk about a trouble to inappropriately positioned matters." In Atkinson, J. M. and Heritage, J. (eds.), *Structures of Social Action: Studies in Conversation Analysis*. Cambridge: Cambridge University Press, pp. 191-222.

Nakajima, S. and Allen, J.F. (1993) "A study on prosody and discourse structure in cooperative dialogues", *Phonetica* 50:197-210.

Wichmann, A. (2000) *Intonation in Text and Discourse*. London: Longman.

Zellers, M. (2009) "Fundamental frequency and discourse meaning in SSBE." Presentation at Phonetics & Phonology in Iberia conference, Universidad de Las Palmas de Gran Canaria, 17-18 June 2009.

Zellers, M., Post, B. and D'Imperio, M. (2009) "Modeling the intonation of topic structure: two approaches." Presentation at 10th Interspeech, Brighton, England, 6-10 September 2009.