Modeling the Intonation of Topic Structure: Two Approaches
Margaret Zellers¹, Brechtje Post¹, Mariapaola D’Imperio²

¹Research Centre for English & Applied Linguistics, University of Cambridge, United Kingdom
²Laboratoire Parole & Langage, Université de Provence, UMR6057 CNRS, France
mkz21@cam.ac.uk, bmbp2@cam.ac.uk, mariapaola.dimperio@lpl-aix.fr

Abstract
Intonational variation is widely regarded as a source of information about the topic structure of spoken discourse. However, many factors other than topic can influence this variation. We compared two models of intonation in terms of their ability to account for these other sources of variation. In dealing with this variation, the models paint different pictures of the intonational correlates of topic.

Index Terms: intonation, discourse, topic structure

1. Introduction
While a wide variety of prosodic research has investigated the discourse functions of intonation on the level of the sentence, especially with regard to the use of focus, there has been relatively little work done on the role of intonation in organizing larger stretches of discourse. The few studies that do exist have proposed that a potential cue to the organization of discourse around a topic is intonation, and specifically the timing of fundamental frequency (F0) peaks relative to the segmental string. The concept of topic has been defined in a variety of ways, but it may be understood here as “aboutness”: what the stretch of discourse is about. Wichmann [1], following Nakajima and Allen [2], proposes that four categories of topic structure can be identified in long discourses, and that these four categories are signaled explicitly by speakers through intonational variation. According to Wichmann’s analysis, these categories signal the degree of newness of information, with the beginning of a topic representing the newest information. Her study of a corpus of BBC newreaders shows late F0 peaks at the beginnings of utterances corresponding with the introduction of new topics, and further timing groups consistent with the other topic structure categories that she identifies.

That study, however, does not take into account research showing that variation in F0 peak timing is a critical cue to the identity of pitch accents (prominence-lending pitch movements) in many languages (cf. Pierrehumbert [3]; D’Imperio [4]; Arvaniti et al [5], inter alia). Even very small changes in the timing of the F0 peak in relation to the segmental stream may be sufficient to cause a shift in categorical perception from, for example, a falling contour (H*L) to a rising contour (L*H). It is essential to consider this issue because the change from a fall to a rise may cause the utterance to have a different meaning by affecting the pragmatic status of the information represented within a pitch-accented phrase (cf. Gussenhoven [6], Pierrehumbert & Hirschberg [7] for theories of intonational meaning). However, this is more complex than it might first seem due to the fact that the first pitch accent in an utterance is likely to be prenuclear rather than nuclear. The nuclear accent is the head of an intonational phrase, and is therefore obligatory in that unit; it is generally perceived as having greater prominence. In English, nuclear accents are always the final pitch accent in the phrase. An intonational phrase may optionally include additional pitch accents preceding the nuclear accent, and these pitch accents are known as prenuclear pitch accents (or (pre-)heads, in the British tradition; cf. Halliday [8], O’Connor & Arnold [9]). Different theories of intonation have proposed a wide variety of methods for distinguishing between these two groups of pitch accents, and there is a lack of consensus as to whether the meaning categories proposed for nuclear accents have the same status in prenuclear accents. Pierrehumbert & Hirschberg [7] as well as Gussenhoven [6] believe that all pitch accents available in nuclear position are also available in prenuclear position, while Ladd and colleagues (in particular Ladd & Schepman [10]) argue that in prenuclear position some accentual distinctions are collapsed. Therefore the predictions regarding the timing of F0 peaks vary widely between these alternatives. Theories that include variation of pitch accent identity in prenuclear position are likely to account for at least part of the F0 timing variation as a cue for distinguishing the identity of the pitch accent, while theories that do not distinguish between prenuclear pitch accents may allow more free variation. Either type of theory must also account for variation contingent on the segmental structure of the utterance.

With regard to Wichmann’s [1] study, then, if we are to accurately identify the prosodic correlates of topic structure, we must first consider which aspects of the phonetics ought to be accounted for by other parts of the intonational theory. The goal of this paper is to evaluate two different models of peak timing variation related to topic structure, and examine some of the implications both for intonational phonology and for discourse prosody.

2. Materials & Methodology
The data for this study were drawn from a production experiment consisting of a read text which was specifically designed to control for segmental influences on intonation as well as having a clearly defined topic structure. Readers of the text were speakers of Standard Southern British English (SSBE), 5 males and 11 females. The recordings were annotated in Praat [11] by one of the researchers (MZ). Pitch and timing data were automatically extracted by Praat. The timing of the F0 peaks was measured in relation to the onset of the vowel of the stressed syllable, and is represented here as a percentage of the vowel length (so the delay value “1” represents an F0 peak that is timed simultaneously with the offset of the stressed vowel) so as to account for variation in speech rate (cf. Silverman & Pierrehumbert [12] for this methodology). The position of the peak is quite variable relative to the stressed syllable; the peak may fall in the postacentual syllable or even later if the stressed syllable as well as the following syllable are short.

Two labeling schemes were applied to the data to represent the two models under examination. All relevant high turning points (or plateau ends, which pattern with high peaks in terms of timing, cf. Knight [13]) which would correspond to the single prenuclear category of (L)H* as per Ladd & Schepman [10] were labeled as H. Then these items were subdivided into groups of H*L or L*(H) on the basis of the analysis provided by Gussenhoven [6] for SSBE. It is important to note that...
although the labeling conventions (i.e. the use of starred tones) in these two methods are similar, the criteria for the analyses are different and therefore result in different groupings.

The data were compared on the basis of four topic structure categories [1,2]. Wichmann [1] specifies that the categories are based on different levels of “newness” of information; however, this is not the only possible interpretation (e.g. Brazil [14] proposes that topic is to do with relative independence of information). The categories used in this study are defined as:

- (New) Topic: a shift in “aboutness” to something different than what precedes it
- Addition: new information on the same topic
- Elaboration: more detail on a previous utterance
- Continuation: completion of information begun in a previous utterance

3. Results and Discussion

3.1. Single pitch accent analysis

When we examine the timing of F0 peaks overall, assuming (as did [1]) that they all belong to the same accentual category, a clear pattern emerges showing new Topics timed latest relative to the stressed vowel, with the other categories following in order: T>A=E>C (F(3, 285)=14.7; p<0.01).

Figure 1: F0 peak timing by topic condition in the single-accent model

This result is comparable to Wichmann’s findings [1], with the exception that it only clearly distinguishes three groups; Additions and Elaborations have slightly different means but the difference is not statistically significant in the current data.

A prediction of the one-pitch-accent analysis is that a large amount of variation in F0 peak timing will be accounted for by the segmental composition of an utterance. The design of the text allows us to control for the identity of the stressed vowel, the specific target word being tested, and the presence or absence of an anacrusis (unstressed syllables preceding the stressed target syllable), among other segmental factors. Therefore it is possible to strip away some of the extraneous variation to look more closely at peak timing variation with relation to topic structure.

First we consider the presence or absence of an anacrusis. Nolan & Farrar [15] have shown that F0 peaks are timed earlier following an anacrusis than if the stressed syllable with which they associate is the first syllable in an utterance. This pattern holds in our data, as seen in Figure 2. Peaks were timed later in the utterance-initial (I) position than in the anacrusis (A) condition in all cases. However, by accounting for some timing variation by including the anacrusis factor, we have lost some of the clear topic-related timing pattern that we found above. Although new Topics still show later-timed peaks than other conditions in Initial position, this pattern does not hold in the Anacrusis condition. Furthermore, within the Initial condition, there are no statistically significant differences between A, E and C, nor are there differences in the anacrusis group (F(3, 118)=0.21, ns).

Figure 2: Timing of F0 peaks in utterances with and without an anacrusis in the single-accent model (anacrusis in left panel)

An additional factor potentially affecting peak timing is the segmental construction of the syllable in question. We can account for this by comparing different occurrences of the same word (and thus the same target syllable) in the data. A pattern similar to that above emerges here, shown in Figure 3.

Figure 3: F0 peak timing within individual words in the single-accent model

Peaks are timed later in the two vowel-initial conditions (Annual and Emory) than in the consonant-initial condition (Rinnering). This is unsurprising given the extra time available for the rise before the onset of the vowel, just as in the anacrusis condition. However, once again we see that although new Topics show later peak alignment in the case of Emory (F(3, 160)=13.51, p<0.01), this is not the case for Annual or Rinnering; and once again, the other topic group distinctions do not attain statistical significance (F(11, 277)=9.41).

Although the patterning of topic structure with F0 peak timing appears clear when all items are combined, when we begin to account for segmental effects, the four- (or three-)
category distinction we originally observed becomes more and more difficult to identify. When we eliminate segmental effects such as the presence or absence of a syllable onset or an anacrusis, the correlation between topic structure and F0 peak timing all but disappears. At best, we are left with an effect of new Topic on the peak timing, which is mitigated if the stressed syllable has a consonant onset or if the utterance has an anacrusis.

3.2. Two pitch accent analysis

An alternative analysis of the pitch accent structure of the data groups the items into two sets, where one is characterized by a rise from the first accented syllable, and the other by a fall. These two groups are differentiated in that the rising category has an additional low tone associated with the stressed syllable; the F0 peak represents a trailing tone following the low starred tone, and is therefore automatically realized later in the word. These groups may also be described tonally as L*H(L) (rising) and H*L (falling); support for this description is given below. In this analysis, we are no longer assuming that F0 peak timing is directly related to the topic structure. Comparing F0 peak timing within the pitch accent groups, we find no statistically significant differences across topic structure categories.

The distribution of the two pitch accents across the topic categories is striking.

3.3. General discussion

The two intonational models presented here provide very different bases for the interpretation of F0 peak timing in relation to topic structure. The one-accent-category model, which would permit peak timing to vary in direct relationship with the topic structure, does indeed show delayed peaks in new Topics. However, once segmental influences on peak timing are accounted for, there is no way to distinguish the other proposed topic structure categories on the basis of peak timing.

In Figure 4, we see that the ratio of L*HL to H*L decreases moving down the topic structure categories: T>>A>E>C (χ²(3, N=289)=60.25, p<0.05). New Topics are more than twice as likely to have an L*HL accent than the other categories.

It is important to note that the distinction between these pitch accent groups is unaffected by the segmental factors, both anacrusis and segmental structure of the target item, that we observed before in the strict timing case. In these conditions, the pattern of the ratio L*HL to H*L remains nearly the same, with T>A>E>C. The only slight exception to this pattern is the anacrusis case, where the A and E groups appear to be reversed in position (see Figures 5 and 6).

The distribution of L*HL and H*L can apparently not be directly tied to topic structure, at least in a categorical sense. If this was the case, we would expect to see (for example) only L*HL in new Topics, and only H*L elsewhere. In the data, though, what we find is that L*HL is simply more likely in new Topics than elsewhere. Gussenhoven [6] has provided an analysis of L*HL which could explain the observed distribution. He proposes a tonal morpheme in English denoting ‘significant information’, realized as the prefixation of an L tone before the pitch accent in question. The new L tone becomes the tone associated with the stressed syllable, altering the pitch accent (by adding an additional meaning element) in the process, as in (1):

\[
L + H^*L \rightarrow L^*HL
\]

We may suggest that information higher in the topic hierarchy, by virtue of being “newer”, would be more likely to be deemed more “significant” than surrounding items, and therefore carry this additional tonal morpheme.

In Figure 5, we see that the ratio of L*HL to H*L decreases moving down the topic structure categories: T>>A>E>C (χ²(3, N=289)=60.25, p<0.05). New Topics are more than twice as likely to have an L*HL accent than the other categories.

It is important to note that the distinction between these pitch accent groups is unaffected by the segmental factors, both anacrusis and segmental structure of the target item, that we observed before in the strict timing case. In these conditions, the pattern of the ratio L*HL to H*L remains nearly the same, with T>A>E>C. The only slight exception to this pattern is the anacrusis case, where the A and E groups appear to be reversed in position (see Figures 5 and 6).

The distribution of L*HL and H*L can apparently not be directly tied to topic structure, at least in a categorical sense. If this was the case, we would expect to see (for example) only L*HL in new Topics, and only H*L elsewhere. In the data, though, what we find is that L*HL is simply more likely in new Topics than elsewhere. Gussenhoven [6] has provided an analysis of L*HL which could explain the observed distribution. He proposes a tonal morpheme in English denoting ‘significant information’, realized as the prefixation of an L tone before the pitch accent in question. The new L tone becomes the tone associated with the stressed syllable, altering the pitch accent (by adding an additional meaning element) in the process, as in (1):

\[
L + H^*L \rightarrow L^*HL
\]

We may suggest that information higher in the topic hierarchy, by virtue of being “newer”, would be more likely to be deemed more “significant” than surrounding items, and therefore carry this additional tonal morpheme.

The two intonational models presented here provide very different bases for the interpretation of F0 peak timing in relation to topic structure. The one-accent-category model, which would permit peak timing to vary in direct relationship with the topic structure, does indeed show delayed peaks in new Topics. However, once segmental influences on peak timing are accounted for, there is no way to distinguish the other proposed topic structure categories on the basis of peak timing.

In Figure 4, we see that the ratio of L*HL to H*L decreases moving down the topic structure categories: T>>A>E>C (χ²(3, N=289)=60.25, p<0.05). New Topics are more than twice as likely to have an L*HL accent than the other categories.

It is important to note that the distinction between these pitch accent groups is unaffected by the segmental factors, both anacrusis and segmental structure of the target item, that we observed before in the strict timing case. In these conditions, the pattern of the ratio L*HL to H*L remains nearly the same, with T>A>E>C. The only slight exception to this pattern is the anacrusis case, where the A and E groups appear to be reversed in position (see Figures 5 and 6).

The distribution of L*HL and H*L can apparently not be directly tied to topic structure, at least in a categorical sense. If this was the case, we would expect to see (for example) only L*HL in new Topics, and only H*L elsewhere. In the data, though, what we find is that L*HL is simply more likely in new Topics than elsewhere. Gussenhoven [6] has provided an analysis of L*HL which could explain the observed distribution. He proposes a tonal morpheme in English denoting ‘significant information’, realized as the prefixation of an L tone before the pitch accent in question. The new L tone becomes the tone associated with the stressed syllable, altering the pitch accent (by adding an additional meaning element) in the process, as in (1):

\[
L + H^*L \rightarrow L^*HL
\]

We may suggest that information higher in the topic hierarchy, by virtue of being “newer”, would be more likely to be deemed more “significant” than surrounding items, and therefore carry this additional tonal morpheme.

The two intonational models presented here provide very different bases for the interpretation of F0 peak timing in relation to topic structure. The one-accent-category model, which would permit peak timing to vary in direct relationship with the topic structure, does indeed show delayed peaks in new Topics. However, once segmental influences on peak timing are accounted for, there is no way to distinguish the other proposed topic structure categories on the basis of peak timing.
structured. A hierarchical analysis following the proposal of Grosz & Sidner [18] might be one alternative.

With regard to the two-accent-category model, we are more free to propose a multiple-level analysis of the topic structure (whether there are three, four, or some other number of levels). However, the two-accent-category data cannot make a definitive statement about the topic structure theory either.

4. Conclusion

By comparing two models of peak timing variation in prenuclear accents, we have been able to evaluate the contribution of a number of additional factors that may influence the prosody of long units of discourse in addition to the topic structure. In particular, with regard to topic structure, anything from the segmental structure of the utterance up to the identity of the pitch accent in question may be a confounding factor in a peak timing analysis, and the intonational model we choose will affect how we deal with these additional factors. Proper models of the prosody of topic structure cannot be made in isolation from models of intonational phonology, but must take into account their predictions. At the same time, modeling topic-structure-related prosodic phenomena can help us to compare, evaluate, and improve current models of intonation.

5. Acknowledgements

Many thanks to Carlos Gussenhoven, Bob Ladd, and Oliver Niebuhr for their helpful comments and suggestions. This research was supported by the EC Marie Curie Training Network “Sound to Sense” (MRTN-CT-2006-035561), and a Programme Blanc grant from the Agence Nationale de la Recherche (Pro-Gram**).

6. References