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 discourse 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.

Figure 1: Accent analyses (schematic). (a) Two-pitch-accent analysis: peak timing causes category shift from fall (H*L) to rise (L*H) (b) One-pitch-accent analysis: all prenuclear accents are high (H* or LH*)

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 based on a read text. The text comprised a fiction narrative which was specifically designed to control for segmental influences on intonation as well as having a clearly defined topic structure. Three categories of target words matched for the identity of the stressed vowel (and where possible using the same word), each three syllables with lexical stress on the first syllable, appeared in either initial or final position within utterances. In initial position in the utterances, the target words could also appear with or without a two-
syllable anacrusis (i.e. the stressed syllable, which was always the first in the target word, could be either the first syllable of the utterance, or the third, following two unstressed syllables). The utterances themselves were varied for topic structure composition on the basis of four topic structure categories [following 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 [11] 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

The topic categories were crossed with the position of the utterance in groups of five utterances, so that each of the categories occurred in each position. The only exception to this crossed design was that, consistent with the way topic is organized in English, new Topics could only occur in the first position in the group, and no other category could occur in that position. This means that to an extent it may be difficult to distinguish between phonetic effects related to the introduction of a new Topic and those related to the beginning of a new group of utterances. In another language which organizes around topic in a different way, it might be possible to separate these effects.

Readers of the text were speakers of Standard Southern British English (SSBE), 5 males and 11 females. The recordings were annotated in Praat [12] 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 proportion 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 [13] for this methodology). The position of the peak is quite variable relative to the stressed syllable; the peak may fall in the postaccentual 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 [14]) 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(L) 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.

### 3. Results and Discussion

#### 3.1. Single pitch accent analysis

When we examine the timing of F0 peaks overall, assuming (as in [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 groups following in order: T>A=E>C (ANOVA: F(3, 285)=14.7; p<0.01). 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.
An additional factor potentially affecting peak timing is the segmental construction of the syllable in question, which we can control for by comparing different occurrences of the same word (and thus the same target syllable) in the data. A pattern similar to that above emerges in this case. Peaks are timed later in the two vowel-initial conditions than in the consonant-initial condition. This is unsurprising given the extra time available for the rise before the onset of the vowel, just as in the anacrusis condition. However, although new Topics once again show later peak alignment in the case of one vowel-initial target word, “Emory” ($F(3, 160)=13.51, p<0.01$), this is not the case for the other two words we tested, and the other topic group distinctions do not attain statistical significance ($F(11, 277)=9.41, ns$).

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 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.

![Figure 4: Pitch accent by topic structure category in the two-accent model](image)

In Figure 4, we see that the ratio of L*HL to H*L decreases moving down the topic structure categories: $T>A=E>C$ ($\chi^2(3, N=289)=60.25, p<0.05$). In fact, 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 the segmental structure of the target item and the presence or absence of an anacrusis, 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 Figure 5).

This is also the only analysis in which L*HL does not outnumber H*L in the new Topics; however, the ratio of L*HL to H*L is still the highest in Topics.

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):

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

‘significant’ + ‘addition’ \rightarrow ‘significant addition’

Affixation of the delay morpheme (following [6])

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.

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.

This leaves the possibility that the categories are primarily signaled by other phonetic factors (for example, the span of the F0 pitch movements; Zellers [16], Zellers & Post [17]). Alternatively, it could mean that the category analysis of topic structure is not the best description of how discourse is 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. This is because the categories themselves do not relate directly to topic, but rather to some other factor that shows a correlation with topic (potentially “significance”, as we have suggested on the basis of Gussenhoven [6], but possibly something else that is as yet unidentified). The distribution of this factor in relation to the topic structure may be categorical or gradient; further research will be required on this question.

Other evidence for the existence of Wichmann’s [1] topic structure categories (e.g. [2, 16, 17]) makes the two-accent-category analysis attractive, because it allows the different topic structure categories to be modeled in intonation. However, it is important to remember that there is no a priori reason to assume that F0 peak timing is actually sensitive to topic structure, especially when we know that it is used to make distinctions on the level of pitch accent identity in nuclear accents. As a consequence, it may be completely irrelevant that there is no consistent variation of F0 peak timing with the topic structure categories in the one-accent-category model; the relevant phonetic cues may be elsewhere.

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 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, as well as to the three anonymous reviewers, 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).