

# *The relationship between accentuation and information status of discourse referents: A corpus-based study\**

DMITRY CITYAEV

---

## **Abstract**

This paper presents empirical results of the corpus-based study on the relationship between accentuation and referent information status. It was found that there is no one-to-one relationship between “new” vs. “given” information and presence vs. absence of accent. Contrary to the widely spread claim that “given” items appear to be deaccented, the results revealed that 79-89% of common nouns and 89-97% of proper nouns which had been previously introduced into the discourse surface as accented. Possible explanations for deaccenting “new” information and accenting “given” information are suggested.

## **1 Introduction**

Prosody can have various functions. Among many of them is signalling “new” information. The relationship between accentuation and information status of discourse referents has been for some time a topic of investigation for both linguists and language engineers. However, not many studies have made a corpus the object of their investigation; yet there has been a growing interest and emphasis on the importance of corpus-based studies in linguistics. This paper sets out to explore the relationship between accentuation and the information status of referents in the discourse using a corpus of recorded speech. The results will be implemented in constructing rules for automatic accent assignment.

It was Halliday (1967a) who drew the attention of scholars in the West to the distinction made by the Prague School linguists between “new information” and “given information”. As he defined it, the primary distinction relates to the state of knowledge which the speaker attributes to the hearer: “new information” is said to be presented by

---

\* This paper forms part of work in progress on the investigation of prosodic features of spoken text. I would like to thank Jill House for most insightful comments on the earlier drafts of this paper.





- (4) Where were your *grandparents* born?  
 {given}

Likewise, “beer” in (3b) in Clark & Haviland’s view (but not in Chafe’s!) is also given information: the information does not necessarily have to be explicitly introduced to be “given”; it can have the status of “given” indirectly via inferencing (“bridging”).

### 3 Taxonomies of “given” and “new” information

**3.0** Below I will survey several proposed taxonomies for “given” and “new” information and propose my own division between “given” and “new” which I will use for examining the relationship between accentuation and referent information status in the discourse.

#### 3.1 Prince (1981)

Prince (1981) suggests a taxonomy for distinguishing NPs expressing new and given information in the discourse which she proposes to call “a taxonomy of the values of Assumed Familiarity” (see Fig. 1). Below is a brief description of the taxonomy with some examples, which are borrowed from Prince.

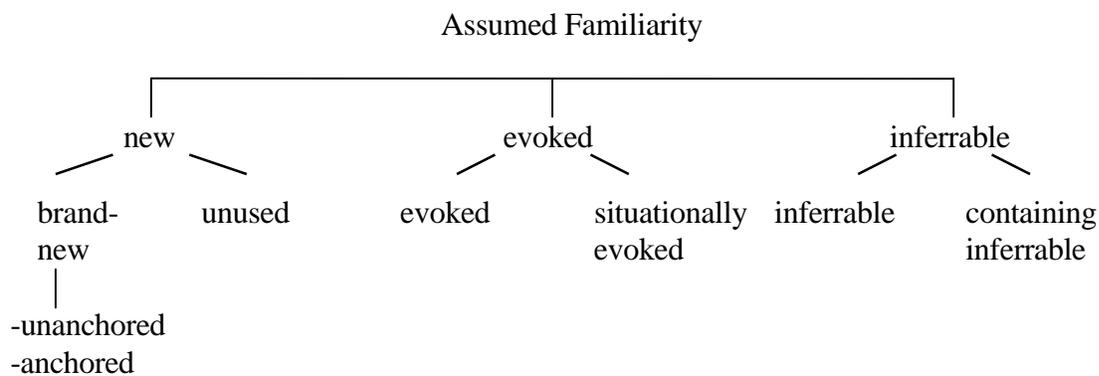


Figure 1. The taxonomy of Assumed Familiarity. Adapted from Prince (1981).

- (5) Pardon, would *you* have a change for a quarter?

- (6) *Noam Chomsky* went to Penn.  
 (7) I got on a *bus* yesterday and the *driver* was drunk.  
 (8) A *guy I work with* says *he* knows your sister.  
 (9) Hey, *one of these eggs* is broken.

Brand-new items are completely new to the hearer. They can be either unanchored, or anchored to some other discourse entity. The NP “bus” in (7) is unanchored brand-new information, whereas “a guy I work with” in (8) is anchored brand-new: it is anchored to the discourse entity “I”. An example of unused type is the NP Noam Chomsky in (6): the knowledge about who this person is is assumed to be shared between the speaker and the hearer.

Evoked entities are those which are already present in the discourse. They can be either previously introduced into the discourse (i.e. anaphoric to their antecedents) as “he” in (8), or they can be situationally evoked as “you” in (5)<sup>2</sup>.

Inferrable items are those which are inferred from the discourse by logical reasoning. For instance, it is known that buses have drivers; thus, “the driver” in (7) is inferrable. Likewise, “one of these eggs” in (9) is said to be a containing inferrable since by virtue of set-member inference, it is inferrable from “these eggs” which in its turn is properly contained within the NP “one of these eggs”.

### 3.2 Brown (1983)

Building on Prince’s taxonomy, Brown (1983) developed her own taxonomy, which includes five categories for entity-referring expressions:

- Brand-new (typically introduced by indefinite NPs)
- New inferred (regularly introduced by definite NPs)
- Evoked context (=Prince’s situationally evoked)
- Evoked current (item which has just been introduced into the discourse and which is currently the entity to which new information is being related)
- Evoked displaced (item which has been introduced into discourse at a point previous to the currently evoked item)<sup>3</sup>

---

<sup>2</sup> Prince notes that situationally evoked entities represent discourse participants and salient features of the extratextual context.

As mentioned earlier, Brown found that the speaker's intonational behaviour accords well with Halliday's descriptions: 87% of brand-new entities and 79% of new inferred entities bear phonological prominence. As far as evoked items are concerned ("given" information), none of the evoked current items are marked by prominence; only 4% of evoked displaced items and 2% of evoked context items are marked by prominence.

### 3.3 Gundel, Hedberg & Zacharski (1993)

Gundel, Hedberg & Zacharski (1993) propose the so-called "Givenness Hierarchy" which in their view reflects the "six implicationally related cognitive statuses":

in focus > activated > familiar > uniquely identifiable > referential > type identifiable

Again, the types of referents are best explained using examples (borrowed from Gundel *et al.*):

- (10) I couldn't sleep last night. *A dog* kept me awake.
- (11) I couldn't sleep last night. *This dog* (next door) kept me awake.
- (12) I couldn't sleep last night. *The dog* (next door) kept me awake.
- (13) I couldn't sleep last night. *That dog* kept me awake.
- (14) I couldn't sleep last night. *That* kept me awake.
- (15) My neighbour's dog bit a girl on the bike. *It* is the same dog that bit me as well.

Type identifiable (10) can be used to refer to a representation of any object. Referentials (11) are used to refer to a particular object; the hearer must either retrieve an existing representation of the intended referent or construct a new representation. Uniquely identifiable referents (12) are associated with definite reference - they are referential and uniquely identifiable with an object. Familiar type (13) is also associated with the ability of the hearer to uniquely identify the intended referent. When the speaker uses the activated type (14), the referent must be represented in the current short-term memory. In-Focus type (15) implies that the referent is not only in the short-term memory but is also in the centre of the current attention.

It is difficult to compare Prince's and Gundel *et al.*'s taxonomies since they take slightly different approaches towards "givenness". One difference, for example, is in

---

<sup>3</sup> Brown adopts Yule's (1981) distinction between "current" and "displaced" evoked items: evoked current items are mostly pronominalised, whereas evoked displaced items are not.

that Gundel *et al.*'s categories "activated" and "in focus" would be collapsed in Prince's taxonomy under "evoked". If we try to sequester the Givenness Hierarchy into "new" and "given" information in the sense of Chafe, then we would probably say that only "activated" and "in focus" type correspond to "given information". The rest is "new" since no matter what cognitive representations of the objects the hearer may have at the time, the referents in (10)-(13) are clearly "new" to the discourse.

### 3.4 Proposed taxonomy of "given" and "new" information

Before proceeding to the investigation of the relationship between accentuation and referent information status, one ought to say a few words about issues one has to face when annotating a text for the information status of its referents. One of them is how many types of referent status we should recognise in our hierarchy. We have seen from the taxonomies discussed above that the number and the criteria on which referents are distinguished vary, depending on one's linguistic assumptions as well as theoretical viewpoint.

Once we have figured out our taxonomy, it seems that the annotation of the text for referent status should proceed without a problem. However, this is not the case. It can be very problematic indeed at times to tell into which of the two or more groups a referent falls, even with very straightforward and well-documented annotation schemes. Poesio & Viera (1998), for example, report on the annotation of definite descriptions for referent information status; their discovery was that inter-annotator agreement was quite low.

It is for these reasons that I have chosen to devise a coarser taxonomy of referent information status than those that have been described above. Since the study is also intended to help develop rules for automatic accent assignment in speech synthesis, it makes sense to draw the parallel between the computer and the speaker. The approach I decided to undertake towards modelling spoken communication is as follows. At the very start, the machine does not know anything about what knowledge is shared between itself (the machine is our prototypical speaker here) and the hearer. It is only as the discourse unfolds that the computer can start treating those entities which it re-introduces as present in the hearer's knowledge/consciousness, and thus attaching a "given information" label to referents.

It appears thus that we can treat the dichotomy "given-new" information as a binary distinction. The referents which are introduced into the discourse for the first time are "new" (even if they happen to be items known to the speaker and the hearer, as in the example "I saw your father yesterday"). Here will belong the following categories

distinguished by Prince: brand-new, unused, and inferrable. Only those entities, which have already been introduced into the discourse (explicitly) or which are anaphorically linked to some referent, will have the status of “given” information. In Prince’s taxonomy, they will belong to evoked entities. In a way, we are treating “givenness” in the sense of Chafe (1976): what the speaker believes to be in the hearer’s consciousness at the time of utterance.

Fig. 2 provides a comparison between the proposed binary distinction of “new” and “given” information and the taxonomies surveyed above.

Proposed distinction	“NEW”	“GIVEN”
Prince (1981)	brand-new, unused inferrable and containing inferrable	evoked and situationally evoked
Brown (1983)	brand-new, new inferred	evoked context, evoked current, evoked displaced
Gundel <i>et al.</i> (1993)	identifiable, referential, uniquely identifiable, familiar	activated, in focus

Figure 2. Proposed taxonomy compared with the other taxonomies.

## 4 Corpus analysis

### 4.1 Corpus description

It was decided to use the PROSICE corpus recorded at UCL for prosody research (Huckvale & Fang, 1996) which contains read speech only. It comprises several monologues which were recorded by a male speaker. The texts used in PROSICE were chosen from the scripted texts in the ICE-GB corpus (Greenbaum, 1996b); they have their origins in the written genre.

The choice of the corpus for this study was primarily guided by the following considerations: high-quality recordings with accurate F0 information (which is important in making decisions about presence/absence of accents), availability of the corpus in the tagged format, distribution through the public domain.

The part of the corpus analysed comprises 45 minutes of recorded speech. There are three texts in it, each totalling about 2000-2500 words. Thus, in total there were about 7000 words analysed in the corpus.

## 4.2 Corpus annotation

The corpus was annotated for part-of-speech, referent information status and absence/presence of pitch accents.

Annotating for presence/absence of pitch accents was quite straightforward. Pitch accents are known to be associated with the stressed syllable of a word (Pierrehumbert, 1980). Thus, one can tell whether the word bears pitch prominence or not via analysing the fundamental frequency contour along with auditory impressionistic listening. Those words which are marked by accents usually have a peak or a valley in the fundamental frequency contour (see Appendix).

Part-of-speech tagging was done automatically and then edited manually. The tagset used for tagging was mainly based on the ICE tagset (Greenbaum & Ni, 1996).

The corpus annotation for "new" and "given" information was first done automatically to save labour: words which were repeated were assigned a "given" information label. During subsequent editing, corrections to the information status were made. When correcting the corpus for referent information status, the following guidelines were borne in mind:

- (i) Compound nouns are normally treated as one referent (e.g. living-room, Baker Street, lamp post).
- (ii) Depending on context, combinations containing two nouns which represent two different referents (especially "of"-phrases) can be "NEW"+"NEW" information, "GIVEN"+"GIVEN", "NEW"+"GIVEN" or "GIVEN"+"NEW" information.

(16) his IMAGES of ARLES  
       {new}            {given}

(17) before the BIRTH of CHRIST  
                   {new}            {new}

- (iii) Although predicating something new about the referent, NPs in apposition are assigned "given" information status, since such parentheticals are anaphoric to the NP introduced just before.



**5.2 Accentuation and “new”/”given” information division**

**5.2.0** Here I will present the main results of the study which reflect how the accentuation of referents interacts with their information status. The interaction was investigated for common and proper nouns, personal pronouns and deictic demonstrative pronouns “this” and “that”. Fig. 3 below is a summary of the results providing data for each text as well as for the whole corpus.

Type of NP	Information Status and Prominence	Text 1	Text 2	Text 3	Whole corpus (all texts)
COMMON NOUNS	new, +acc	95%	96%	93%	95%
	new, -acc	5%	4%	7%	5%
	given,+acc	79%	86%	89%	86%
	given, -acc	21%	14%	11%	14%
PROPER NOUNS	new, +acc	97%	98%	100%	98%
	new, -acc	3%	2%	0%	2%
	given,+acc	89%	91%	97%	94%
	given, -acc	11%	9%	3%	6%
PERSONAL PRONOUNS	given,+acc	11%	5%	8%	9%
	given, -acc	89%	95%	92%	91%
PRONOUNS “THIS”, “THAT”	given,+acc	70%	100%	45%	67%
	given, -acc	30%	0%	55%	33%

Figure 3. Proportions of the referents with different information status as a function of their accentuation, for each text separately and for the whole corpus.

**5.2.1 Accentuation and “new”/”given” information status of referents expressed by common nouns and proper nouns.** As far as common nouns are concerned, 95% of 1120 referents which enter the discourse as “new information” are accented, which leaves the remaining 5% of the referents with “new information” status unaccented. Referents which represent “given information” tend to display quite an unexpected behaviour: 86% out of 355 referents are in fact accented, which clearly contradicts the well-known claim that “given” information is likely to be deaccented: only 14% of the referents with “old information” status surface as unaccented.

With the referents expressed by proper nouns, the situation is quite similar. 98% of 122 proper nouns which were labelled as “new information” are marked by accent (leaving 2% unaccented); 94% of 147 proper nouns which were re-introduced in the

discourse (= “given” information) are accented (which is higher than 86% of accented, “given information” referents expressed by common nouns).

Some examples of those nouns which are accented on their repeated mention are presented below in Fig. 4.

<i>Word</i>	<i>repeated</i>	<i>accented</i>	<i>Word</i>	<i>repeated</i>	<i>accented</i>
Provence	5	5	Pope	4	4
France	3	3	Sarah	9	9
Les Arlescans	3	3	gypsies	3	3
Rome	4	4	bridge	5	4
Arles	4	3	saints	3	3
Avignon	7	7	Miss Crompton	11	10
Romans	5	5	William	34	30

Figure 4. Some referents with the number of the times they have been repeated in the text and the number of the times they have been accented on the repeated mention.

It appears that there is no one-to-one relationship between the “new”/“given” information and presence/absence of accent. We have seen that referents bearing “given” information status are very often accented. Also, we have found that some of the referents (even though a very small percent) bearing “new” information status are de-accented. If we examine these more closely, we can come up with hypotheses as to why “new” items may be deaccented.

The de-accenting of “new” information can be caused by the so-called “Rhythm Rule”, proposed in Liberman & Prince (1977), and elaborated further on in Selkirk (1984). Selkirk argues that some constituents surface as de-accented (or bear only secondary prominence) to avoid rhythmical clashes in the metrical grid. Here are some examples from our corpus:

(21) Miss CROMPTON’S brother, JACK  
       {given}                    {new} {given}

(22) BEFORE the birth of CHRIST  
       {new}                    {new}



All these hypotheses require further investigation, of course.

One should also wonder why “given” information items are accented on so many occasions. It is quite possible that the treatment of nouns in apposition as “given” information contributed to the increase in the percentage of “given” information being accented. However, I would not expect this to be the most crucial factor. A possible explanation for the phenomenon of accenting “given” information can be adherence to rhythmicality (to avoid rhythmical lapses in the metrical grid – see Selkirk, 1984). In text 27, for instance, the PP “before the birth of Christ” was mentioned twice and on the second occasion, both the referents (“birth” and “Christ”) were still accented. Even though they represent “given” information, the speaker chose to place accents on them – most probably to satisfy the well-formedness constraints on metricality.

Accenting “given” information may have something to do with topicality. Take, for instance, the referent “William” which is by far the most often repeated one in the corpus. It is mentioned 35 times throughout the text and on 30 occasions of its repeated mention, it is accented. One may wish to argue that “William” is a topic and for this reason it is accented. Yet, topics represent “given” information and on those grounds would be expected to surface as deaccented. Indeed, on most repeated mentions, “William” appears in the subject position, which makes it a topic. Thus, it may as well be that topics tend to be accented rather than deaccented (for similar thoughts, see Nooteboom & Kruyt, 1987).

Let us compare the results of our study with those obtained by Brown. As far as “new items” are concerned, the results are quite comparable: they tend to bear phonological prominence. However, “given” items behave differently in our study than in Brown's. In the PROSICE, 86% of the common nouns and 94% of the proper nouns, representing “given” information, are accented. According to Brown's findings, only 4% of the referents of the corresponding referent type (“evoked displaced” items) were found to bear accents. Such a polarity can be possibly explained by the different discourses investigated. I shall come back to the discussion of this question in the last section.

**5.2.2 Pronouns and accentuation.** It is also interesting to look at the behaviour of personal pronouns with respect to their accentuation. As already mentioned above, personal pronouns are anaphors and thus following our criteria stand for “given” information. We would expect personal pronouns to be deaccented in most cases for two reasons: (i) they represent “given” information and (ii) they belong to the “function” word-class group.

In reality, the situation is as follows. It is true that most of the personal pronouns in the corpus appear to be deaccented. However, 9% out of total 317 do bear accents. Some of the examples are below:

(32) They offered it (= job) to someone else, but HE changed his mind so they had to make do with ME.

(33) THEY were mostly in their fifties and sixties. I [accented] was thirty seven.

The phenomenon of accenting pronouns which refer to entities either introduced previously in the discourse (e.g. “he”) or introduced situationally (e.g. “you”, “I”) can be possibly explained within the centering framework elaborated on in Grosz, Joshi & Weinstein (1995). According to Grosz *et al.*, if a referent is in the subject position, then it is more likely to be the highest-ranked “forward-looking centre”. In (33), for example, the forward-looking centre in the first sentence is “they” (it is also a “backward-looking centre” since it refers to “colleagues” - a referent introduced previously). Yet, in the second sentence there is a different referent in the subject position - “I”: there occurred a centre shift: the referent which is the backward-looking centre in the second sentence (“I”) is not the same as the referent which is the backward-looking centre in the first sentence (“they”). Centre shifting in a way disrupts local coherence and increases processing load. Thus, the speaker needs to attract a special attention of the hearer that there occurred a centre shift and for this reason he probably chooses to place an accent on the pronoun.

**5.2.3 Deictic demonstratives and accentuation.** Deictic demonstrative pronouns “this” and “that” which stand for NPs also deserve special attention. Being NPs, they fall in Gundel *et al.*’s taxonomy under the referent type “activated”. Although “this” and “that” are clearly anaphoric and thus represent “given” information, we would, however, because of their deictic properties (see below), expect them to bear phonological prominence. The empirical results support our hypothesis: 67% of 27 deictic demonstrative pronouns were found to be accented. Altenberg’s findings are very similar: 75% of all the demonstrative pronouns expressed by “that” bear phonological prominence. Consider the two examples from our corpus below:

(34) We ALL know where THAT can lead.

(35) But THIS DEMONSTRATES CONCLUSIVELY that....

Accenting the demonstrative pronouns “this” and “that” may have something to do with the fact that they are deictic, i.e. “pointing”, especially to actions: in (34) “that” refers to the “goings-on under the bridge” and in (35) “this” refers to William and the Outlaws’ “deciding to live like St Francis and the Franciscans”. Again, one may resort to the centering framework for possible explanations: a pronoun “this” or “that” is accented since we need to make it a “backward-looking centre” in the absence of a concrete referent prior to uttering “this”/“that”.

## 6 Conclusion

This study has been concerned with the investigation of the relationship between referent information status and accentuation. Most of the results obtained in this empirical study support claims made previously in the literature. They are:

- function words tend to be deaccented and content words tend to be accented (Hirschberg, 1993) (informal observations indicate that of the function words, it is quantifiers and numerals which tend to be accented on most occasions)
- personal pronouns tend to be deaccented; yet, when accented, they signal centre shifts from one referent to another (Nakatani, 1998), or contrastiveness
- “new” information is most likely to be accented (Brown, 1983)

It was also found that deictic demonstrative pronouns which stand for NPs tend to be accented.

In general, I have found that there is no one-to-one relationship between accentuation and the information status of referents. In the type of discourse that I have analysed (monologue, read speech), referents bearing “new” and “given” information happened to be both accented and deaccented. The most striking observation (which does not normally find enough acknowledgement in the literature!) was that there is quite a high percentage of referents representing “given” information that do bear phonological prominence. Thus, the widely spread generalisation that “given” information tends to be deaccented is too broad.

The comparison between the results of our study and the results obtained by Brown revealed differences as far as the behaviour of “given” information is concerned. In our corpus, common nouns and proper nouns which represent “given” information tend to be accented. Brown’s NPs of that type (=“evoked displaced” items) are almost always de-accented. As suggested in the previous section, such a difference can arise because

of the different discourses investigated. I investigated monologue, read speech, where the distance between co-referring referents can be large. Brown looked primarily at short, task-oriented dialogues – an example of discourse where distances between “given” information and previous mention of the referent are short.

Brown & Yule (1983: 168) note (referring to Brown’s study) that

the conversations produced by the speakers were relatively short... and the entities they were concerned with were very limited.

Thus, Brown & Yule conclude that

it is hardly surprising that the few entities, established in such short conversations, should be expected by the speaker to remain accessible to the hearer.

They further hypothesise that “in longer conversations, speakers may feel they have to reinstate previously mentioned information”. This echoes with Chafe’s discussion about consciousness: “One indisputable property of consciousness is that its capacity is very limited”, he writes. “As new ideas come into it, old ones leave” (Chafe, 1976: 32). If, following Chafe, our consciousness has a limited capacity, then it will be able to keep only a certain number of referents at one particular time. It becomes clear now why some referents which are mentioned again (“given” information) may surface as accented: the speaker assumes that these referents must have left the addressee’s consciousness by the time they are re-introduced into the discourse.

The interaction between accentuation and referents with “given” information status may be better explained if one considers the number of sentences separating the mention of the referent from its last mention, or the number of referents intervening between the mention of the referent and its last mention. In this respect, it would be interesting to see how the model of the discourse structure presented in Grosz & Sidner (1986) can help us explain the relationship between presence/absence of accent and “given”/“new” information status. We may find that with every new paragraph, the focus stack containing all the referents introduced in the previous paragraph is popped, and subsequently all the referents which were “given” information are treated again as “new” information. It is only in the neighbouring two-three sentences (belonging to one and the same paragraph) that we may expect a re-introduced referent carrying “given” information status to be always deaccented. This will be the future direction for this

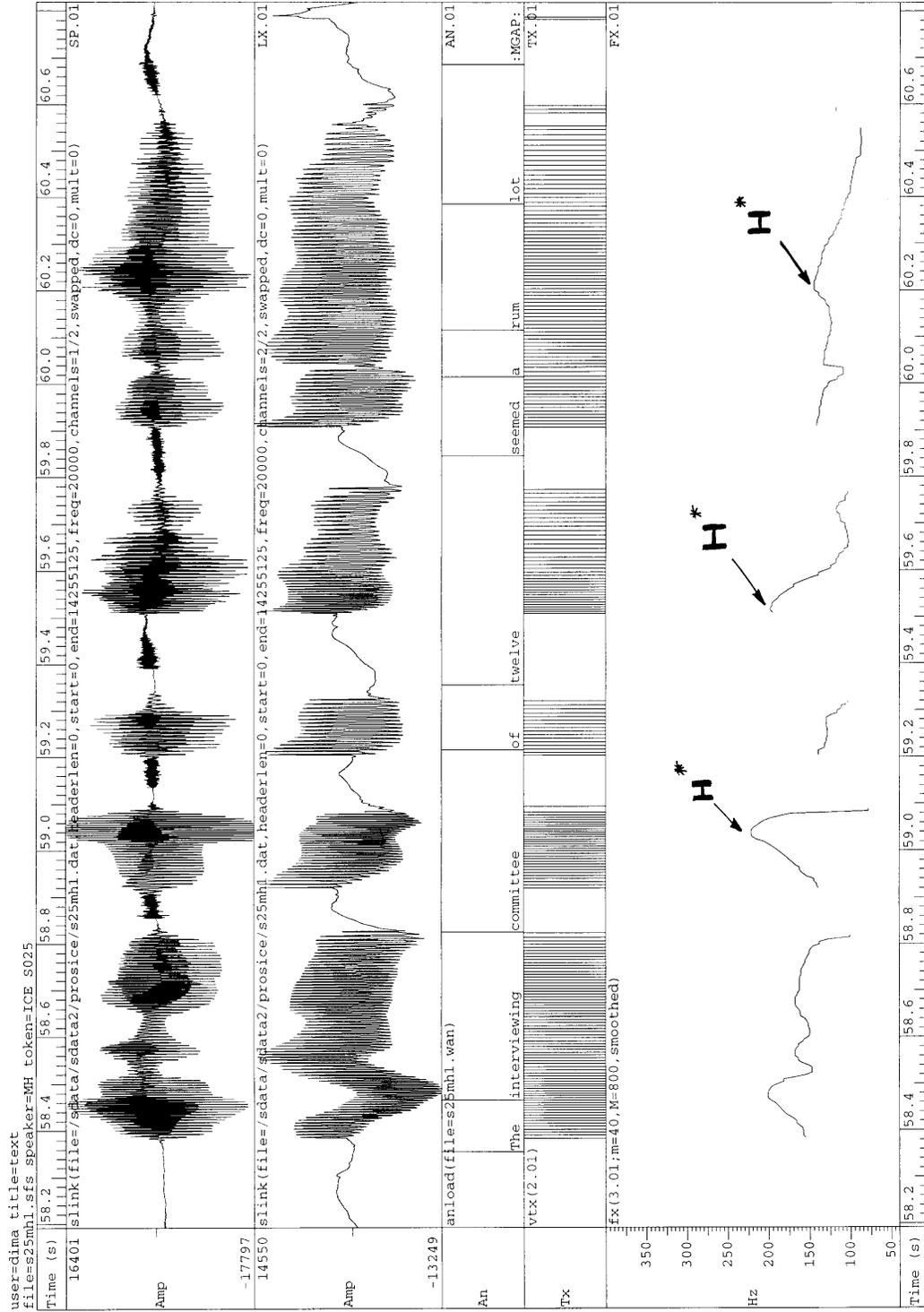
research where a discourse model will be implemented to the PROSICE corpus and the results will be viewed in the light of this model.

## References

- Altenberg, B. (1987) Prosodic patterns in spoken English. *Lund Studies in English* 76.
- Brown, G. (1983) Prosodic structure and the given/new distinction. In Cutler, A. & R. Ladd (eds.). *Prosody: Models and Measurements*. 67-77. Berlin: Springer Verlag.
- Brown, G. & G. Yule (1983) *Discourse Analysis*. Cambridge: Cambridge University Press.
- Clark, H. & S. Haviland (1977). Comprehension and the given-new contrast. In Freedle, R. (ed.). *Discourse Production and Comprehension*. 1-40. Hillsdale: Lawrence Erlbaum Associates.
- Chafe, W. (1974) Language and consciousness. *Language* 50: 111-133.
- Chafe, W. (1976) Givenness, contrastiveness, definiteness, subjects, topics and point of view. In Li, C. (ed.). *Subject and Topic*. New York: Academic Press.
- Cruttenden, A. (1997) *Intonation*. Cambridge: Cambridge University Press.
- Greenbaum, S. (1996a) *The Oxford English Grammar*. New York: Oxford University Press.
- Greenbaum, S. (ed.) (1996b) *Comparing English Worldwide: The International Corpus of English*. Oxford: Oxford University Press.
- Greenbaum, S. & Y. Ni (1996). About the ICE Tagset. In Greenbaum, S. (1996b) 92-109.
- Grosz, B. & C. Sidner (1986) Attention, intentions and the structure of discourse. *Computational Linguistics* 12(3): 175-204.
- Grosz, B., A. Joshi & S. Weinstein (1995) Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* 21(2): 203-225.
- Gundel, J., N. Hedberg & R. Zacharski (1993) Cognitive status and the form of referring expressions in discourse. *Language* 69: 274-307.
- Gussenhoven, C. (1984) *On the Grammar and Semantics of Sentence Accents*. Dordrecht: Foris.
- Halliday, M. A. K. (1967a). Notes on transitivity and theme in English (part 2). *Journal of Linguistics* 3: 199-244.
- Halliday, M. A. K. (1967b). *Intonation and Grammar in British English*. The Hague: Mouton.
- Halliday, M. A. K. & R. Hassan (1976) *Cohesion in English*. London: Longman.
- Hirschberg, J. (1993) Pitch accent in context: Predicting intonational prominence from text. *Artificial Intelligence* 63: 305-340.
- Hirschberg, J. (1999) Communication and prosody: function aspects of prosody. *Proceedings of ESCA International Workshop on Dialogue and Prosody*. 7-15. Veldhoven: ESCA.
- Huckvale, M. & Fang, A. (1996). PROSICE: a spoken English database for prosody research. In Greenbaum, S. (1996b). 262-279.
- Kuno, S. (1972) Function sentence perspective. *Linguistic Inquiry* 3: 269-320.
- Kuno, S. (1978) Generative discourse analysis in America. In Dressler, W. (ed.). *Current Trends in Textlinguistics*. 275-294. Berlin: de Gruyter.
- Leech, G., M. Deuchar & R. Hoogenraad (1982) *English Grammar Today: A New Introduction*. London: Macmillan.

- Lieberman, M. & A. Prince (1977) On stress and linguistic rhythm. *Linguistic Inquiry* 8: 249-336.
- Nakatani, C. (1998) Constituent-based accent prediction. *Proceedings of ACL/COLING*. . 939-945. Montreal: ACL.
- Nooteboom, S & J. Kruyt (1987) Accents, focus distribution and the perceived distribution of given and new information: an experiment. *Journal of Acoustical Society of America* 82: 1512-1524.
- Pierrehumbert, J. (1980) *The Phonology and Phonetics of English Intonation*. PhD dissertation. Distributed by Indiana University Linguistics Club.
- Poesio, M. & R. Viera (1998) A corpus-based investigation of definite description use. *Computational Linguistics* 24: 183-216.
- Prince, E. (1981) Toward a taxonomy of given/new information. In Cole, P. (ed.). *Radical Pragmatics*. New York, Academic Press.
- Selkirk, E. (1984) *Phonology and Syntax: The Relation between Sound and Structure*. Cambridge: MIT Press.
- Selkirk, E. (1995) Sentence prosody: intonation, stress and phrasing. In Goldsmith, J. (ed.). *Handbook of Phonological Theory*. Oxford: Basil Blackwell.
- Yule, G. (1981) New, current and displaced entity reference. *Lingua* 55: 41-52.

Appendix: Example of the analysed PROSICE corpus



An example of the annotated corpus. The first window displays the waveform; the orthographic transcription is in the third window; the bottom window contains the fundamental frequency contour. The sentence is: "The interviewing COMMITTEE of TWELVE seemed a RUM lot". Accents are marked as "H\*".