

PROMINENCE JUDGEMENTS AND TEXTUAL STRUCTURE IN DISCOURSE¹

Monique E. van Donzel and Florian J. Koopmans-van Beinum

Abstract

In this paper the perceptual evaluation of a method to analyse the information structure of a discourse (Van Donzel, 1994) is discussed. In this evaluation listeners were presented with spoken versions of a retold story in Dutch, and were asked to mark those parts in the text they perceived as being emphasized, under the assumption that these are the most salient parts of the discourse. The results of the experiment indicate that the method represents fairly well the way listeners perceive information structure, but that it needs to be refined and adapted in certain aspects. These adjustments are discussed as well.

1 Introduction

When listening to a spoken discourse, listeners have certain ideas about the structure of the incoming text. They perceive certain words or word groups as more prominent than others, while they are also able to detect different types of boundaries, such as sentence boundaries and paragraph boundaries (e.g. Swerts, 1994; Blaauw, 1995).

Much work is done to investigate the relation between discourse structures and their prosodic features, using independent frameworks of discourse structure (e.g. Geluykens & Swerts, 1994; Grosz & Hirschberg, 1992; Grosz & Sidner, 1986; Hirschberg & Grosz, 1992; Nakatani, 1995; Nakatani *et al.*, 1995; Swerts & Geluykens, 1994; Terken & Hirschberg, 1994). These investigations are primarily concerned with the overall global structure of whole discourses. Hirschberg & Grosz (1992), however, analyse their discourses both at a global level (the structure of the discourse constituents which form the whole discourse) and at a local level (parentheticals, quotations, tags, and indirect reported speech). In this paper we will concentrate on the relation between the perceived prominence and the internal 'focal' structure of a text, thus on the (local) utterance level rather than on the (global) paragraph or discourse level. Our local level, however, differs from the one used by Hirschberg & Grosz (1992): by 'local' we mean the structure of the different types of information within the utterance.

In Van Donzel (1994) we developed an independent framework for discourse analysis, in which the internal focal structure of a text is based on pragmatic theories about discourse structure (Mann & Thompson, 1988; Chafe, 1987; and Prince, 1981), rather than on acoustic features such as intonation and accent peaks. By using this

¹ Parts of this paper were presented at the XIIIth International Congress of Phonetic Sciences, Stockholm, 13-19 August 1995, and also appeared in the Proceedings of that Congress (Van Donzel & Koopmans-van Beinum, 1995).

framework we can avoid the circularity of acoustic features being already included in the definition of focus. This method applies to both written texts and verbatim transcriptions of spontaneously produced spoken texts. The analysis thus obtained reflects the structure of a text, based on the written text alone.

The organisation of the paper is as follows. First of all, we will briefly discuss different approaches to analyse the informational structure of texts (section 2). Then we will present a perception experiment (section 3). Section 4 discusses the implications, and the conclusions are presented in section 5.

2 Approaches to the analysis of information structure

2.1 Introduction

Basically there are three approaches to analyse the information structure of a sentence or a text with respect to prominence. This is also referred to as the 'accent placement debate' (cf. Baart, 1987):

1. *syntactic* approach: 'within this approach, researchers attempt to establish a direct relation between the lexico-syntactic structure of a sentence and its accentuation, and to express this relation in a formal way' (op. cit., p. 10).
2. *pragmatic-semantic* approach: 'according to this view, accent placement is the result of an interplay between various factors, among which the semantic content of a sentence and the relation of a sentence to its context are the most important' (id., p. 11).
3. '*focus*' approach: 'within this framework, it is attempted to overcome the problems of both the syntactic and the semantic-pragmatic approaches by positing two stages in the derivation of an accent pattern. In the first stage, a speaker selects one or more constituents of a sentence as material to be emphasized, or focussed upon. The outcome of this stage [...] is unpredictable in principle, since a speaker is free to decide which distribution of focus best suits his communicative intentions. In the second stage, however, the exact locations of the accents within focussed constituents are automatically derived on the basis of lexico-syntactic structure' (id., p. 11).

Both the syntactic and the 'focus' approach assume highly structured material, such as question/answer pairs, produced in so-called laboratory speech, and would thus not seem very suitable to be applied to spontaneous speech. In our view the pragmatic approach is more suitable to analyse discourses and spontaneous speech, since this one takes into account the notions 'knowledge of the world' and 'context', which are crucial in spontaneous speech.

Another way of deciding on the accent placement, is the use of text-to-speech systems, for instance PROS (Dirksen & Quené, 1993; Quené & Kager, 1993). These kinds of systems use algorithms that automatically analyse a given text into different parts conveying different kinds of information, and then assigning pitch accents and other prosodic features to the most salient parts. However, at this stage, for texts to be correctly analysed, they usually need to consist of grammatical and correctly structured sentences. Thus, this method would not be useful to analyse 'real' spontaneous speech, since this type of speech often shows a lot of hesitations and unfinished ('ungrammatical') utterances.

These considerations, together with the observed circularity (the definition of focus already includes acoustic features) led us to develop an 'objective' method to analyse

discourses and spontaneous speech, independently of acoustic features, and taking into account pragmatic aspects crucial to freely retold speech. In a next step, the material can be analysed acoustically, to investigate the acoustic-phonetic correlates of the objective information structure.

2.2 Proposed discourse analysis

In this section we will briefly present the method that we used to analyse the informational structure of discourses. A more elaborate discussion can be found in Van Donzel (1994).

First of all, the verbatim transcription of a spontaneous discourse is divided in clauses or utterances. A clause is defined as a group of words or concepts taken together on functional or semantic grounds. This means that a clause can consist of a verb plus nouns and determiners or modifiers, but also of an idiomatic expression. In general, these clauses are rather small. They can be compared to what Chafe (1987; 1994) calls 'intonation units', but our clauses are evidently not based on any intonational features, to avoid the circularity mentioned in the introduction.

Nominal constituents can be classified as follows, using so-called 'textual labels'. A *brand new* (bn) element refers to information that is completely new in the addressee's context. This addressee can be either the reader of the transcription or the listener of the spoken version of the discourse. Brand new elements are usually indefinite nouns or generic expressions. An *unused* (u) element is also new, but the listener can place the information it expresses directly in his/her discourse model. These are usually definite nouns or proper names. An element is labeled as *inferrable* (i) if the speaker assumes that the listener can infer it from the preceding context or from his/her knowledge of the world. *Evoked* elements have already been mentioned in the discourse. They can be I) *textually evoked* (et): the noun is evoked by a real pronoun, II) *displaced textually evoked* (etd): the noun cannot be evoked by a pronoun because the referent is too far back in the discourse, the full noun is used to avoid ambiguity, III) *situationally evoked* (es): the referent of a noun or pronoun can only be found in an extra textual context. *Modifiers* (mod) express some kind of degree or quality. *Orientations* (or) express temporal or locational orientations at the beginning of clauses.

The method just described is strongly oriented towards the receiver of the message, contrary to the speaker oriented 'focus' approach mentioned in section 2.1. Therefore, our method is very suitable to be related to listeners' judgements.

In Van Donzel (1994) a description is given of a highly detailed analysis, in which all nominal constituents of a text are labeled according to the status of the information they express. At that point, however, the method does not fully account for the fact that all verbs, and not just the ones expressing new information, can also be labeled according to their information status. So, presently, verbs are classified using the labels *unused*, *inferrable* and *evoked* in the same way as for nominal constituents. Since verbs can intuitively not be *brand new*, the label *unused* is proposed. Evidently no distinction can be made in *textually* or *situationally evoked* verbs, so they will be classified as *evoked*. The verb phrase as a whole is labeled, the auxiliary and the main verb thus are considered as a unitary concept. Prepositions which are part of a verb are related to them by giving an index to both of them (see also Van Donzel & Koopmans-van Beinum, 1995).

3 Listening experiment

3.1 Introduction

A listening experiment was carried out to investigate whether there is perceptual evidence for the proposed analysis, in other words, what is the relation between the perceptual judgements of listeners and the textual analysis? What 'extra' cues can prosody add to the textual structure of a discourse?

3.2 Material

A short story in Dutch (*Een triomf* by S. Carmiggelt, 1966) was read aloud after some preparation time by four male and four female native speakers of Dutch (read version). All speakers were students or staff members of the Institute of Phonetic Sciences. After a short break they were asked to tell this story in their own words, with as many details as possible (retold version). During the retelling a listener was present in the recording room, to create a more natural telling situation. This retold version was literally transcribed by the first author, including all hesitations and false starts, but without any punctuation mark or capital. The next day this verbatim transcription was read aloud by the same speaker (re-read version). The speaker was encouraged to read the text carefully before reading it aloud, to mark punctuations, and to correct hesitations or false starts. This was explicitly *not* done by the transcriber, since any change or mark in the text will influence the way the speaker reads the text. Thus, the retold version and the re-read version are in principle at least lexically identical. All recordings were made in a sound treated room on DAT-tape. The speakers participated on a voluntary basis.

3.3 Procedure

3.3.1 Textual analysis and hypotheses

The informational structure of the eight transcribed retold versions was evaluated by the first author. These analyses were presented to a panel of five text analysts, all familiar with discourse theories. The proposed text analyses were discussed, and this resulted in an ultimate convention for labeling. Where necessary the originally proposed analyses were adapted.

We hypothesize that all information that is in some way *new* to the discourse, will be perceptually judged as prominent (labels *bn, u, i, mod*), while the information *already mentioned* will not be perceptually judged as prominent (labels *or, et, etd, es*). In this respect we will compare the two speaking styles retold and re-read, as well as possible differences between male and female speakers.

3.3.2 Perceptual evaluation

Since we are mainly interested in the perceived structure of spontaneous speech, we only used the retold and re-read versions in this experiment. This resulted in 16 different texts (8 speakers x 2 versions). These were randomly ordered, and presented to 16 listeners, in such a way that every listener evaluated 3 different texts and that each text was scored by 3 different listeners. All listeners were students or staff members of the University of Amsterdam. Student listeners were paid for their participation. The

spoken versions were presented over headphones, the verbatim transcription of the spoken text was used as an answer sheet.

The listeners were instructed to evaluate the spoken versions in terms of prominence, using only the speech signal. Each listener was presented with an individual tape containing four different spoken versions of the story, either a retold version or a re-read transcription, from four different speakers. The first text was used as an exercise. They were asked to underline those parts in the text they perceived as being emphasized by the speaker, on the basis of the speech sound only, so explicitly not on the basis of the written text, and then to judge the relative prominence of these parts on a scale from 1 (very emphasized) to 3 (less emphasized). These marks, however, do not necessarily represent the linguistic terms of primary, secondary, and ternary stress. The two hours limit given to fulfill the task was sufficient for all listeners.

3.4 Results

3.4.1 Overall judgements

To get a first impression of how the perceptual judgements of prominence might be related to the textual analyses, we normalized the data to percentages and summed all judgements (Table 1). The three perceptually most relevant labels are *unused* (22%), *brand new* (17%) and *modifier* (16%). This is as can be expected since these labels represent words containing 'new' information. Thus, 55% of all underlined parts were 'new' in the discourse. Chi-square tests on the absolute numbers reveal that there are significantly more 'new' parts perceived as emphasized than 'inferred' or 'evoked' parts ($\chi^2=48.4$, $df=1$, $p<.001$).

When looking at the remaining prominence judgements, we find the following: *evoked textually* (8%), *evoked textually displaced* (14%) and *evoked situationally* (1%), all labels referring to 'given' information. Again, these relatively low percentages, apart from *etd*, can be expected, since evoked items will generally not be pronounced with much emphasis. However, *the evoked textually displaced* items seem to be perceived as more emphasized than other evoked items. This is not surprising either, since it is exactly these items that cannot be pronominalized, they have to be 'refreshed', and thus are 'new' in a certain sense. For example, 'the forest' is referred to at a later point in a discourse about a walk in the woods, not by means of the pronoun 'it' but by repeating the full noun 'the forest' to avoid ambiguity.

The *inferred* items represent information that is neither completely new nor completely evoked. From the parts perceived as emphasized, 14% is inferred. This might suggest that this category is indeed a valid one in the analysis. The 'rest' group (7%) consists of the items *orientation* (or) and zero judgements (0). Zero judgements can be either textual labels without a specific prominence judgement (horizontally in Table 1), or concepts judged as perceptually prominent which did not have a textual label (vertically in Table 1).

When looking at the relative prominence judgements (1, 2, or 3), we find that 28% of all items are judged with a 1, 45% with a 2, 27% with a 3 and 0,3% did not have a specific perceptual judgement level. This indicates that listeners did use the whole scale of possibilities.

This first look at the data suggests that there does seem to exist a relation between the textual analysis and the overall prominence judgements of listeners. Elements that add new information to the discourse are perceived as emphasized more often than elements representing information that is already evoked earlier in the discourse. Information that can be inferred from other elements in the discourse is also perceived as emphasized in

a number of cases. However, listeners do not seem to give a particular judgement (1, 2, or 3) to a particular textual label (*or*, *mod*, *bn*, etc.); so there does not seem to be a clear correlation between a certain judgement level and a certain textual label. In almost half of the cases listeners judged a 2 (this is done significantly more often than the other two judgements: $\chi^2=35.6$, $df=1$, $p<.001$), which may indicate that only in extreme cases a 1 or a 3 was judged. Therefore, in the rest of this paper we will take into account only the total percentage of prominence judgements per textual label, as given in the right-most column of Table 1.

Table 1. Perceptual prominence judgement matched against textual label, normalized to percentages, for all speakers and all listeners together. Totals are given for each label separately as well as for the category as a whole. Prominence levels 0, 1, 2, and 3 refer to the relative prominence judged by the listeners; the textual labels 0, or, mod, bn, u, i, etd, ed, and es refer to the different categories of information from the discourse analysis, as explained in section 2.2.

judged prominence	0 = no level	1 = very emphasized	2 = less emphasized	3 = little emphasized	total	
textual label						
0	0.04	1.50	2.16	1.81	5.52	7.34
or	0.00	0.51	0.96	0.34	1.82	
mod	0.06	4.34	6.96	4.25	15.61	55.47
bn	0.00	5.14	7.69	4.55	17.37	
u	0.08	6.43	9.95	6.03	22.49	
i	0.04	3.97	6.27	3.73	14.01	14.01
etd	0.05	4.07	6.39	3.81	14.32	23.18
et	0.04	1.59	3.95	2.59	8.16	
es	0.00	0.24	0.20	0.26	0.70	
total	0.31	27.78	44.54	27.36	100%	100%

3.4.2. Speaking styles and sexes

In this section we will look at possible differences between prominence perception in the two speaking styles retold and re-read, and between the ways in which the discourses produced by male and female speakers are perceived.

The first two columns of Table 2 present the overall percentage of prominence judgements, for the retold and re-read speaking styles. There do not seem to be very large differences between the two styles; they differ at most 2 absolute percentage points, the effect of speaking style is not significant ($\chi^2=19.8$, $df=8$, $p=.01$). We expected larger differences between the two speaking styles, since they are perceptually quite distinct. This was stated in a small classification experiment, in which 8 male and 8 female students and staff members of the Institute volunteered. The listeners were asked to classify 1,5 minute medial fragments from the spoken texts as either 'spontaneous' or 'read', and to mark the degree in which they were sure of their choice (1 for 'not sure', to 5 for 'very sure'). They answered correctly in 90% of the cases, and were in 60% of the cases 'very sure' of their choice. This suggests that the listeners in the evaluation experiment, who listened to the entire spoken text, were able to hear the difference in speaking style as well.

We do find, however, that whenever the retold speaking style dominates in number of prominence judgements, this is exactly for the major categories from Table 1 (*brand new*, *unused*, *inferable* and *evoked textually displaced*). This might follow from the

fact that the method of text analysis is developed from discourse theories based on spontaneous speech.

The last two columns present the overall percentage of judgements, for the male and female speakers separately. In some cases, the male and female speakers seemed to behave differently. As for the major categories, the male speakers got higher scores than the female speakers. The female speakers, however, emphasized much more *modifiers* than did the male speakers. This might suggest that the female speakers had a more elaborate way of telling, while the male speakers were more 'compact'. The effect of sex of speaker is significant ($\chi^2=32.7$, $df=8$, $p=.0001$).

Table 2. Overall percentage prominence judgements, broken down for speaking style and sexe of speaker. (revised data, cf. Van Donzel & Koopmans-van Beinum, 1995)

		retold	re-read	male	female
'discourse markers'	0	4.6	6.6	5.4	5.9
	or	1.4	2.2	1.8	1.8
'new'	mod	15.6	17.1	13.2	19.3
	bn	17.1	15.8	18.3	14.6
	u	23.0	22.0	23.4	21.7
'inferrable'	i	14.4	13.4	14.7	13.1
'evoked'	etd	15.1	13.2	14.5	13.7
	et	8.1	9.1	8.3	9.0
	es	0.7	0.7	0.6	0.8
	total	100%	100%	100%	100%

3.4.3 Zero judgements

Finally, something has to be said about the so-called 'zero judgements'. Overall, they cover about 6% of all labels, meaning that 6% of the items underlined by the listeners as being prominent, did not have a textual label in the original text analysis, and thus they could not be classified. At a closer look, these items appeared to be mainly discourse markers (*well, thus, so, etc.*) or discourse connectives (*and, or, etc.*). (The label *orientation* can in fact also be seen as a discourse marker.) However, cases in which an auxiliary was perceived as emphasized without the main verb being perceived as such, fall in this category zero judgements as well. This does not mean that auxiliaries should be labeled separately, since they form a unitary concept with the main verb (cf. Chafe, 1987). Normally, the main verb should be the accentable part of the unity. However, the auxiliary (as any word in the text) can be emphasized for contrastive reasons. Contrastive prominence is considered here as a separate class of information within the analysis. The informational status of a concepts can be changed or altered for contrastive reasons; for example, the pronoun 'he', which normally represents *evoked* information, can become *new* or *inferrable* by adding contrastive prominence to it.

3.5 Preliminary conclusions and re-analysis

In this section we will discuss some preliminary conclusions concerning the data presented above. The method to analyse the textual structure of a discourse by means of nine textual labels, used in the previous sections, gives a rather detailed analysis. The notions defined in this method are in fact too complex to be used as working definitions of types of information status. Therefore, we want to reduce the possible classes of

information to the four most important categories, namely *new*, *inferrable*, *evoked* and *discourse markers*. This division can easily be made, and follows almost naturally from the Tables 1 and 2. As can be seen, the labels *mod*, *bn* and *u* behave in an identical way. These three can thus form the category new. The same goes for the labels *et* and *es*: these two form the category evoked. The *zero judgements* and the *orientations* can be grouped together also, since both are in fact one and the same category of 'markers' indicating the major parts of a discourse. So, these will constitute the category discourse markers. Both Tables 1 and 2 show that the labels *i* and *etd* are perceived alike. This is not surprising when we realize that the *evoked displaced* items really are not 'evoked' in the same way as are the textually and situationally evoked ones. Since their referent is too far back in the discourse, they can at a certain point not be referred to by means of a pronoun. The full noun is used to avoid ambiguity, and thus it can be expected that such a noun is emphasized by the speaker. Thus, both *inferrables* and *evoked displaced* items can be grouped together in the category inferrable.

It is precisely here where we see a clear interaction between the pragmatic structure of a text and the perceptual prominence judgements, or, in other words, where prosody does not coincide with textual structure. On the basis of the textual analysis, the *evoked displaced* items would fall in the category evoked. But, prosody decides otherwise, since these items are clearly perceived as salient. This fact is also observed by Prince (1992) on non-prosodic grounds, namely on the basis of the outcome of a computer programme called VARBRUL, which performs binomial logit analyses on linguistic data. She groups together inferrables with Discourse-old Nonpronominals (our *textually evoked displaced* items).

The eight transcriptions of the retold version were re-analysed according to the adjustments mentioned above. Where necessary, the analyses of information status were adapted. This did not affect the analysis in a very significant way.

To sum up, we present the four major categories of information status, with the labels included in each of them, as well as the type of lay-out used to mark those categories. An example of this re-analysis is given in Figure 1 below, contrastive information is marked in capitals:

- | | | |
|----|---------------------|--|
| 1. | new | brand new, unused, modifiers |
| 2. | <i>inferrable</i> | inferrable, evoked textually displaced |
| 3. | <u>evoked</u> | evoked textually, evoked situationally |
| 4. | [discourse markers] | orientations, zero judgements |
| | CONTRASTIVE | parts of the discourse conveying contrastive information |

<p><u>het</u> eeh gaat₁ over₁ twee mensen <u>die</u> wonen in de stad ‘it is about two people who live in the city’ en [op een morgen] worden₁ <u>ze</u> wakker₁ ‘and one morning they wake up’ en [dan] zien <u>ze</u> <u>dat</u> het heel hard gesneeuwd heeft ‘and then they see that it has snowed very heavily’ <u>het</u> is [dus] een verhaal in de winter ‘so it is a story in wintertime’ en <u>ze</u> besluiten om die dag eens in het bos te gaan kijken ‘and they decide to go and see in the woods’ hoe <u>het</u> er [dan] <u>daar</u> uit ziet how it looks overthere’ de stad UIT het bos IN # ‘out of the city, into the woods’ in het bos is <u>het</u> eeh heel heel dik <u>besneeuwd</u> ‘in the woods there is a whole lot of snow’ de takken van de jonge bomen <u>die</u> buigen₁ over₁ ‘the branches of the trees bend over’ en <u>daar</u> moeten <u>ze</u> soms onderdoor₁ kruipen₁ ... ‘and sometimes they have to crawl underneath them’ ...</p>

Figure 1. Example of a text analysis, using only four categories of information structure, expressed in different kinds of lay-out. English translations are given after each utterance between quotes.

4 Implications

At this point we also would like to answer the question posed in section 3.3.1: ‘what is the relation between judgements of perceived prominence and the textual information structure?’ However, before we can say anything about the relation between prosody and textual structures, we will have to take a closer look at the internal structure of the text material. This will also enable us to say something about the differences between speakers, since they all produced texts of different lengths, and since no two speakers retell the story in the same way.

Table 3 presents the total number of items (‘concepts’) for each of the four major categories, based on the textual analysis, and the total length of each text (in number of words) for the eight speakers, where text refers to the verbatim transcription of the retold version. Furthermore, the mean percentages of items in each category in the original written text are presented, as well as the mean percentages ‘perceived as emphasized’ in each category.

Evidently, the length of the text is different for the eight speakers. However, when we compare the distribution of each category as a function of the total length of the text, we see that it is comparable for all speakers (see Table 3): 43% of each text consists of *new* concepts, 20% is *inferable*, 33% is *evoked*, and 4% is *discourse marker*. This group of discourse markers, however, includes only those discourse markers as defined in section 2.2 (clause initial orientations). Zero judgements are thus *not* included, since we are dealing with the textual distribution, and not with the one assigned by the listeners. This is, for comparison with the original text not very important, since very little clause internal discourse markers occur in the original text. The fact that the distributions are alike indicates that apparently all speakers produced similar textual discourse structures (there is no speaker effect: $\chi^2=23.1$, $df=21$, $p=.34$).

It is interesting to compare the distribution of categories in the spontaneously retold versions to the distribution in the original written text as given by the author. The total length of the original text is 615 words. The internal distribution of the original text is the following: 51% *new* concepts, 13% *inferrables*, 29% *evoked* items and 7% *discourse markers*. This distribution does not differ very much from the one found in the spontaneous versions. This could mean that the distribution is perhaps 'universal' and applies to all discourses. Whether this is the case remains to be seen. As far as we know, nothing is said about this matter in the literature. We find, however, that there are more *new* concepts and *discourse markers* in the original written text, and less *inferrable* and *evoked* items than in the retold version (the effect of version is, however, not significant for the two texts: $c^2=7.9$, $df=3$, $p=.047$). This could be an effect of 'style': in written text there is no real interaction between the producer of the text and the one perceiving it, so the writer has to make the structure of the text as clear as possible: the reader cannot immediately intervene to ask for clarifications. The most important concepts are the *new* ones, of which there are significantly more in the original version, and the *discourse markers*, since these ones mark the beginnings of different parts of the text.

Let us now turn to our hypotheses about the relation between information structure and perceived prominence. We expected 'new' material (i.e. categories *new* and *inferrable*) to be perceived as emphasized, and 'given' material (i.e. categories *evoked* and *discourse markers*) not to be perceived as emphasized. When we look at the bottom part of Table 3, we find that only 52% of the *new* and 58% of the *inferrables* are perceived as being emphasized, while still 13% of the *evoked* items and 18% of the *discourse markers* are perceived as being emphasized. This was not what we expected on the basis of our hypotheses. Our next step will be to relate the structural analyses of the texts to the judgements of perceived emphasis, to see whether items that occupy certain positions in the utterance are more likely to be perceived as being emphasized than items in other positions. This can, however, not directly be derived from the results of the experiment described in this paper, since not every text is perceptually evaluated by the same listeners. Additional tests are currently being carried out in which all listeners in a group evaluate all eight texts.

Furthermore, these results are rather surprising compared to other findings in the literature, for instance Brown (1983). Using a taxonomy based on Prince (1981), she finds that 87% of all *new* concepts, 79% of the *inferrable* concepts and only 4% of the *evoked* concepts are accented. However, her material consisted of highly structured instruction dialogues, and she did not include verbs or discourse markers in the labeling. These differences could very well explain the varying results.

Our results deviate from other findings in the literature (cf. also Nootboom & Kruyt, 1987), which can probably be explained by differences in speech material used. We think, however, that we have gained some insight in the relation between the pragmatic structure of a text and the prominence judgements of that same text assigned perceptually.

Table 3. Total number of concepts for each category (percentages given between brackets) and total length of text (in number of words), for the eight speakers; mean percentage concepts of each category in spontaneously retold as well as in original written version; mean percentage of concepts judged as emphasized in each category for the spontaneously retold version.

		new	inferrable	evoked	discourse markers	length
s	1	107 (42)	51 (20)	83 (33)	12 (5)	539 wds
p	2	100 (42)	38 (16)	93 (39)	6 (3)	460 wds
e	3	118 (45)	53 (20)	78 (30)	11 (5)	582 wds
a	4	91 (39)	50 (21)	79 (34)	16 (6)	504 wds
k	5	97 (44)	45 (20)	71 (32)	8 (4)	490 wds
e	6	67 (45)	37 (25)	41 (28)	4 (2)	361 wds
r	7	92 (43)	39 (18)	76 (36)	5 (2)	415 wds
s	8	117 (49)	40 (17)	73 (30)	11 (4)	511 wds
mean % concepts of each category in <i>spontaneously</i> retold version						
		43	20	33	4	100%
% concepts of each category in <i>original written</i> version						
		51	13	29	7	100%
mean % <i>emphasized</i> in each category in spontaneously retold version						
		52	58	13	18	

5 Conclusions

The aim of the perception experiment described above was to investigate the relation between judgements of perceived prominence on the one hand, and textual structure labels, based on a prosody-independent method on the other hand. The results suggest that the relation is roughly as can be expected: *new* items are perceived as emphasized more often than are *given* items. In the literature we already find that there is no one-to-one relation between new/given information and plus/minus accented, or plus/minus prominent. In our spontaneous material, we find that the relation between information structure and perceived prominence is even less one-to-one as could be concluded from the literature (cf. Nooteboom & Kruyt, 1987). We found that only 52% of all *new* items in a discourse is perceived as being emphasized, while still 13% of all *given* items is perceived as being emphasized. For the *new* items this is rather a low percentage, for the *given* items it is rather high, both cases surprisingly deviate from data found in the literature so far. This could be an effect of differences in material between the experiments so far and our experiment (laboratory speech read aloud vs. a spontaneously retold story).

The next step in our project will be to investigate structurally, acoustically, and perceptually, on the one hand the remaining *new* and *inferrable* items not perceived as emphasized, and on the other hand the *given* items perceived as emphasized. This could possibly explain these patterns.

The method we used to analyse our material results in an internal 'focal' structure of a discourse, in which the major categories of information type are specified. We see these different parts of information as potential landing sites for focal accents. The exact place of these accents, i.e. the constituent placed 'in focus', remains to be determined, possibly through a structural analysis of the different parts of the discourses. It can very well be the case that only *new*, *inferrable* or *evoked* concepts occupying a certain position within the clause are marked as prominent. Ayers *et al.* (1995) mention in this respect the downstepping of accents in spontaneous speech: after the focal accent in the

utterance, the remaining non-focal accents are downstepped, even if the accents correspond to material not previously mentioned in the discourse, i.e. *new* material. This could also very well explain the low percentage *new* concepts perceived as being emphasized. A close examination of this downstepping will be necessary as well.

Acknowledgements

The authors would like to thank Rob van Son for his help in processing the data, and for making it possible to run the classification experiment on the net. We thank Louis Pols for careful reading of the draft.

References

- Ayers, G., G. Bruce, B. Granström, K. Gustafson, M. Horne, D. House & P. Touati (1995) 'Modelling intonation in dialogue', in: K. Elenius & P. Branderud (Eds) *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, volume 2, p. 278-281.
- Baart, J.L.G. (1987) *Focus, syntax, and accent placement. Towards a rule system for the derivation of pitch accent patterns in Dutch as spoken by humans and machines*, Doctoral dissertation, Leiden University.
- Blaauw, E. (1995) *On the perceptual classification of spontaneous and read speech*, Doctoral dissertation, Utrecht University.
- Carmiggelt, S. (1966) 'Een triomf', in: *Fluiten in het donker*, ABC Boeken, Amsterdam.
- Chafe, W.L. (1987) 'Cognitive constraints of information flow', in: R.S. Tomlin (Ed.) *Coherence and grounding in discourse*, Typological studies in language 11, John Benjamins Publishing Company, Amsterdam/Philadelphia, p. 21-51.
- Chafe, W.L. (1994) *Discourse, consciousness, and time. The flow and displacement of conscious experience in speaking and writing*, The University of Chicago Press, Chicago & London.
- Dirksen, A. & H. Quené (1993) 'Prosodic analysis: The next generation', in: V.J. van Heuven & L.C.W. Pols (Eds) *Analysis and synthesis of speech*, Speech Research 11, Mouton de Gruyter, Berlin, p. 131-146.
- Geluykens, R. & M.G.J. Swerts (1994) 'Prosodic cues to discourse boundaries in experimental dialogues', *Speech Communication* 15, p. 69-77.
- Grosz, B.J. & J. Hirschberg (1992) 'Some intonational characteristics of discourse structure', in: J.J. Ohala et al. (Eds) *Proceedings of ICSLP 92*, Banff, p. 429-432.
- Grosz, B.J. & C.L. Sidner (1986) 'Attention, intentions and the structure of discourse', *Computational Linguistics* 12 (3), p. 175-204.
- Hirschberg, J. & B.J. Grosz (1992) 'Intonational features of local and global discourse structure', *Proceedings of the DARPA workshop on spoken language systems*, Arden House, p. 441-446.
- Mann, W.C. & S.A. Thompson (1988) 'Rhetorical Structure Theory: Toward a functional theory of text organization', *Text* 8 (3), p. 243-281.
- Mann, W.C. & S.A. Thompson (1992) *Discourse description: Diverse linguistic analyses of a fund-raising text*, Benjamins, Amsterdam.
- Nakatani, C.H. (1995) 'Discourse structural constraints on accent in narrative', submitted to *Progress in speech synthesis*.
- Nakatani, C.H., J. Hirschberg & B.J. Grosz (1995) 'Discourse structure in spoken language: Studies on speech corpora', *AAAI spring symposium on empirical methods in discourse interpretation and generation*, Stanford.
- Nooteboom, S.G. & J.G. Kruyt (1987) 'Accents, focus distribution, and the perceived distribution of given and new information: An experiment'. *Journal of the Acoustical Society of America* 82 (5), p. 1512-1524.
- Prince, E.F. (1981) 'Toward a taxonomy of Given-New information', in: P. Cole (Ed.) *Radical Pragmatics*, Academic Press, New York, p. 223-255.
- Prince, E.F. (1992) 'The ZPG Letter: subjects, definiteness, and information-status', in: W.C. Mann & S.A. Thompson (Eds) *Discourse description: Diverse linguistic analyses of a fund-raising text*, Benjamins, Amsterdam, p. 295-325.

- Quené, H. & R. Kager (1993) 'Prosodic sentence analysis without parsing', in: V.J. van Heuven & L.C.W. Pols (Eds) *Analysis and synthesis of speech*, Speech Research 11, Mouton de Gruyter, Berlin, p. 115-130.
- Swerts, M.G.J. (1994) *Prosodic features of discourse units*, Doctoral dissertation, TU Eindhoven.
- Swerts, M.G.J. & R. Geluykens (1994) 'Prosody as a marker of information flow in spoken discourse', *Language and Speech* 37 (1), p. 21-43.
- Terken, J.M.B. & J. Hirschberg (1994) 'Deaccentuation of words representing 'given' information: effects of persistence of grammatical function and surface position', *Language and Speech* 37 (2), p. 125-145.
- Van Donzel, M.E. (1994) 'How to specify focus without using acoustic features', *Proceedings of the Institute of Phonetic Sciences* 18, University of Amsterdam, p. 1-17.
- Van Donzel, M.E. & F.J. Koopmans-van Beinum (1995) 'Evaluation of discourse structure on the basis of written vs. spoken material', in: K. Elenius & P. Branderud (Eds) *Proceedings of the XIIIth International Congress of Phonetic Sciences*, Stockholm, volume 3, p. 258-261.