EVALUATION OF DISCOURSE STRUCTURE ON THE BASIS OF WRITTEN VS. SPOKEN MATERIAL

Monique E. van Donzel & Florien J. Koopmans-van Beinum
Institute of Phonetic Sciences/IFOTT, University of Amsterdam, The Netherlands

ABSTRACT

In this paper we present the results of an experiment in which firstly we asked text analysts to evaluate the verbatim transcriptions of a retold story in terms of 'informational structure', using a method [2] based on linguistic knowledge and intuition. We then had listeners underline emphasized words and scale them for degree of emphasis in the spoken versions of the same story, but on the basis of the speech sound only. The prediction was that in the latter case linguistic knowledge may be overruled by the actual speech sound. Results show that this indeed seems to be the case.

INTRODUCTION

The structure of information in *written* texts usually becomes clear by the use of typographic means. In *spoken* texts it is generally assumed that the speaker may use various acoustic means to assign structure, for instance by accenting important words. In written texts words can also be perceived as being more or less important, in this case there is evidently no relation with accents.

In the often used elicitation method of *question/answer pairs* the informational structure (focus distribution) can be described using the labels 'new' vs. 'old' information, where 'new' usually refers to 'accented' and 'old' to 'not accented'. Focus is thus defined through intonation. However, this kind of definition may lead to circularity in that the possible acoustic features are already included in the definition itself.

How the focal structure of a whole *discourse* should be traced is less clear. Therefore, we developed a method [2] using various theories about discourse structure, in which the focal structure of a text is based on the informational structure rather than on the acoustic features, thus avoiding the circularity mentioned above.

The goal of our experiment was to see if there is a relation between the informational structure, based on linguistic knowledge, and prominence judgements of listeners based on the speech sound. Possible differences between different speaking styles and between sexes are discussed as well.

METHODS

Speakers, text analysts, and listeners

Four male and four female speakers, all native speakers of Dutch, were selected as speakers for the experiment. They were all students or staff members of the Institute of Phonetic Sciences. Five text analysts, all familiar and experienced with text analyses, participated in the evaluation of the written text. The speakers and text analysts participated on a voluntary basis. Seven male and nine female students and staff members of the University of Amsterdam, all native speakers of Dutch, participated as listeners in the experiment. The student listeners were paid for their participation.

Stimuli and recordings

The speakers were asked to read aloud a short story in Dutch (*Een triomf* by Simon Carmiggelt). After a short break they were asked to tell the same story in their own words, as detailed as possible (the 'retold' version). During the retelling of the story, a listener was present in the recording room, to create a natural telling situation. From this retold version a verbatim transcription was made by the first author, and the speaker was asked to read aloud this transcription the next day (the 're-read' version). All recordings were made in a sound treated room on DAT-tape.

Method of text analysis

In this section we will briefly present the method used to analyse the informational structure of the recorded discourses. This method is a combination of several theories about the structure of discourses [1, 3, 4]. Because of space limitations, we will discuss here only the labels at the word level.

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 listener's context. This usually regards 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. This 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.

Verbs are classified using the labels *unused*, *inferrable* and *evoked* in the same way as for nominal constituents. The verb phrase as a whole is labeled, the auxiliary and the main verb 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.

Written evaluation

The informational structure ('focal structure') of the transcribed retold versions of the four male and four female speakers was evaluated using the method described in section 2.3. The analyses were made 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 proposed analyses were adapted. An example of parts of one of the texts and its analysis is presented in Table 1.

Table 1. Example of a text analysis.

het [es] eeh gaat [u] over twee mensen [bn] die wonen [u] in de stad [u]
en op een morgen [or] worden$_1$ ze [et] wakker$_1$ [u]
en dan [or] zien [u] ze [et] dat het heel hard [mod] gesneeuwd [u] heeft [i]
het [es] is dus een verhaal [bn] in de winter [i] [ai]
en ze [et] besluiten [u] om die dag [i] eens in het bos [u] te gaan kijken [u]
hoe het [et] er dan daar [et] uit ziet [i]
de stad [etd] uit het bos [etd] in
in het bos [etd] is het eeh heel heel dik [mod] besneeuwd [e]
de takken van de jonge bomen [i] die buigen$_1$ [u] over$_1$
en daar [et] moeten ze [et] soms [mod] onderdoor$_1$ kruipen$_1$ [u] ...

Perceptual evaluation

The 16 listeners were instructed to evaluate the retold versions and the re-read versions in terms of prominence, using only the speech signal which was presented over headphones. Each listener was presented with an individual tape containing four different spoken versions of the story (the first text was used as an exercise), either a retold version or a re-read transcription, from four different speakers. They were asked to underline the parts of the discourse 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 do however not necessarily represent the linguistic terms of primary, secondary and ternary stress. The verbatim transcription of the spoken text was used as an answer sheet. There was a two hour time limit to the task.

RESULTS

Textual structure and perceptual prominence

Each text was evaluated by three different listeners. For each of the eight verbatim transcriptions the analysis based on the text alone was taken as reference point. The perceptual judgements were compared to these analyses. For every text, style and listener a confusion matrix was made, in which the labels from the text analysis were matched against the

prominence judgements 1, 2 and 3. 'Zero labels' (0) were added to cover the cases in which a word was underlined but no judgement was given (zero perception) and the ones in which a word was underlined that did not have a proper label in the text analysis (zero text analysis). This resulted in 48 matrices (8 speakers x 2 styles x 3 listeners per text).

Overall matrix

To get a first impression of how the textual analysis might be related to the perceptual analysis, we normalized to percentages and summed all 48 matrices (Table 2). 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' items in the discourse ($p \leq 0.001$, df=1, χ^2=48.4).

When looking at labels referring to 'given' information, we find the following: *evoked textually* (8%), *evoked textually displaced* (14%) and *evoked situationally* (1%). 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 the discourse not by means of the pronoun 'it' but by repeating the full noun 'the forest' to avoid ambiguity.

Table 2. Overall matrix, normalized.

	0	1	2	3	total
or	0.00	0.51	0.96	0.34	1.82
mod	0.06	4.34	6.96	4.25	15.61
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
et	0.04	1.59	3.95	2.59	8.16
etd	0.05	4.07	6.39	3.81	14.32
es	0.00	0.24	0.20	0.26	0.70
0	0.04	1.50	2.16	1.81	5.52
total	0.31	27.78	44.54	27.36	100%

The *inferrable* items represent information that is neither completely new nor completely evoked. From the parts perceived as emphasized, 14% is inferrable. This might suggest that this category is indeed a valid one in the analysis. The 'rest' group (7%) consisted of the items *orientation* (or) and zero judgements (0).

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 perceptual judgement. 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 and a certain textual label. In almost half of the cases listeners judged a 2 ($p \leq 0.001$, df=1, χ^2=35.6), 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 judgements.

Differences between speaking styles and between sexes

In this section we will look at possible differences between the two speaking styles, and between the ways in which male and female speakers are perceived.

The first two columns of Table 3 present the overall percentage of 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%, these effects do not appear to be significant. We expected larger differences between the two speaking styles, since they are perceptually quite distinct. However, whenever the retold speaking style dominates, this is exactly for the major categories from Table 2 (*brand new, unused, inferrable* 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 behaved differently. As for the major categories, the male speakers scored higher 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'.

Table 3. Overall percentage judgements, broken down for retold/re-read speaking style and for male/female speaker.

	retold	re-read	male	female
or	1.44	2.19	1.77	1.84
mod	14.34	16.89	12.94	19.53
bn	18.36	16.38	18.82	14.65
u	23.06	21.91	23.25	21.29
i	14.89	13.14	14.64	12.80
et	7.39	8.93	7.79	9.20
etd	15.32	13.33	15.05	13.63
es	0.68	0.73	0.71	0.83
0	4.52	6.51	5.04	6.25
total	100%	100%	100%	100%

Finally, something has to be said about the so-called 'zero judgements'. Overall, they cover about 5% of all labels, meaning that 5% of the words underlined by the listeners did not have a textual label or no judgement was given, and thus could not be classified. At a closer look, these words appeared to be mainly discourse markers (*well, thus, so,* etc.) or discourse connectives (*and, or,* etc.). However, cases in which an auxiliary was perceived as emphasized without the main verb being perceived as such, fall in this category as well.

DISCUSSION

In this section we will try to test our hypothesis that linguistic knowledge may be overruled by the actual speech sound in assigning structure to spoken texts.

The data show clear evidence for the three major categories *new, inferrable* and *evoked.* New words are expected to be perceived as being emphasized. Inferrable and evoked words, however, are not expected to be perceived as being emphasized so often, since these words represent information that is known at some level.

When looking at our results, we find that in exactly these cases there is a difference between the expected data and the observed data, especially when regarding the *inferrable* and the *evoked textually displaced* items: these are perceived as emphasized quite often. This indicates that in these cases, the actual speech sound does overrule linguistic knowledge, since emphasis is not expected.

Furthermore, the method of text analysis will need to be extended to *discourse markers,* to account for a part of the zero judgements, and to so-called *'contrastive accents'* to account for the occurrence of, among other things, emphasized auxiliaries.

ACKNOWLEDGEMENTS

The authors would like to thank Rob van Son for his help in processing the matrices, and Louis Pols for comments on the paper in general.

REFERENCES

[1] Chafe, W. (1987) 'Cognitive constraints on information flow', in: R.S. Tomlin (Ed.) *Coherence and grounding in discourse*, John Benjamins, Amsterdam/ Philadelphia, pp. 21-51.

[2] Donzel, M.E. van (1994) 'How to specify focus without using acoustic features', *Proceedings* **18**, Institute of Phonetic Sciences, University of Amsterdam, pp. 1-17.

[3] Mann, W.C. & S.A. Thompson (1988) 'Rhetorical Structure Theory: Toward a functional theory of text organisation', *Text* **8** (3), pp. 243-281.

[4] Prince, E. (1981) 'Toward a taxonomy of Given-New information', in: P. Cole (Ed.) *Radical Pragmatics*, Academic Press, New York, pp. 223-255.