Perception of discourse boundaries in spontaneous speech in Dutch

van Donzel, M. E. and Koopmans-van Beinum, F. J.

journal: TMH-QPSR
volume: 37
number: 2
year: 1996
pages: 105-108

http://www.speech.kth.se/qpsr
Perception of discourse boundaries in spontaneous speech in Dutch

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

Abstract
This paper describes a perception experiment in which listeners were asked to mark various discourse structures in the verbatim transcription of a spontaneously retold story in Dutch, while listening to the spoken version of the story. They marked perceived discourse boundaries by means of conventional punctuation marks. Previously, the transcriptions had been analysed for discourse structure, on independent, non-prosodic grounds.

The aim of the experiment was to see in what way the perceived boundaries coincide with the objectively determined ones, and in what way listeners agree on a certain type of boundary, and in a later stage, to investigate by means of which cues listeners could make their specific decisions.

Introduction
It is generally assumed that in spoken discourse, speakers may use various acoustic means to assign structure to the text, for instance by marking highly important words as more prominent, and by chunking the text into smaller pieces. Listeners are also able to detect these structures: they usually have ideas about the structure of the incoming text, in terms of phrasing and prominence (e.g. Lehiste, 1979; Swerts, 1994).

Within spoken texts, as used in normal, communicative situations, speakers introduce concepts at various points in the discourse. A concept introduced earlier, can be referred to by the speaker at later points in the discourse, for instance by the use of a pronoun (the given/new distinction, cf. Prince, 1981). Listeners usually can trace the referent of such a pronoun without much difficulty. This means that they keep track of the information status of objects, expressed in earlier stages of the discourse. This also means that it is necessary to do so, otherwise, it would not be possible to process the pronouns in a meaningful way.

At some points in a discourse, however, pronominalization is not an available option to refer to earlier introduced items. When the referent is too far back in the discourse, the use of a pronoun may lead to ambiguity. In such cases, the speaker needs to repeat the full noun. Thus, apparently, there has been some kind of discourse boundary, across which pronominalization is not possible (cf. also Horne et al, 1993). This indicates that listeners also keep track of more global structures, such as the division in terms of paragraphs (cf. Chafe, 1994), since a reference is made to information 'far away' in the discourse.

Listeners are thus able to tell when new discourse units begin, and when they end. This has convincingly been shown, for structured discourse tasks, by e.g. Swerts (1994) and Blaauw (1995).

Prosodically marked boundaries
The prosodic organization of texts read aloud has widely been investigated (see for instance Bruce, 1982). Boundaries can be prosodically marked by the speaker with for instance a pause, with a boundary-marking pitch movement or with both. Other possible cues are preboundary lengthening, variation in intensity and in voice quality. Full prosodic boundaries include both melodic and pausal cues, while less heavy boundaries are marked with only a pitch movement or a pause. In Dutch, melodic means to mark boundaries include rising and level tunes, generally associated with non-finality or continuation, and falling tunes, associated with finality (e.g. Blaauw, 1995). This leads to the general expectation that in spoken discourse and text read aloud, phrases (noun phrases with determiners and modifiers) and clauses (noun phrases grouped together on semantic or functional grounds) are marked as non-final, while paragraphs (clauses grouped together as dealing with the same topic) are marked as final by the speaker. This also enables the
listener to keep track of the global structure of the discourse.

Spontaneous discourse, even if it is less strictly structured, and characteristically contains more disfluencies and hesitations, can be structured along the same lines as more carefully structured (or prepared) material. The same global build up into phrases, clauses, and utterances or paragraphs applies. In spontaneous discourse as well, listeners keep track of the information status of concepts introduced earlier (cf. the frequent use of pronouns). For more details about the structuring of spontaneous discourse see, for instance, Chafe, 1994; Nakatani et al, 1995.

The goal of our experiment was to see how discourse structure, and more specifically discourse boundaries, are perceived in spontaneous speech in Dutch. Assuming that speakers mark the different parts of the discourse acoustically, listeners are able to infer from these acoustic aspects the intended structure. The listener needs this information to process the discourse in a coherent and meaningful way, and to know where to expect new discourse units. We will thus focus on global discourse features.

Methods

Speakers, stimuli, and recordings

Eight native speakers of Dutch, four male and four female, were selected. They were all students or staff members of the Institute of Phonetic Sciences of the University of Amsterdam.

The speakers were asked to read aloud a short story in Dutch ('A Triumph' by Simon Carmiggelt). After a short break, they were asked to retell the same story in their own words, with as many details as possible (the 'retold version'). During the retelling a listener was present in the recording room to create a more natural telling situation. All recordings were made in a sound-treated room on DAT-tape.

Discourse analysis

An independent framework for discourse analysis was developed earlier (Van Donzel, 1994; Van Donzel and Koopmans-Van Beinum, 1995a, b), in which the information structure (focal structure) of a discourse is based on pragmatic grounds, rather than on acoustic features. The application of this method results in the division of a discourse in terms of clauses and paragraphs on a global level, and on a local level in terms of information structure, using the four major categories new, inferrable, evoked and discourse marker ('so', 'then' etc.). For each phrase (or 'concept') in the discourse the information status is determined. A clause is defined as phrases taken together on functional or semantic grounds; this coincides with a division for written text in main and subordinate clauses. A new paragraph occurred whenever the speaker started a new (sub)topic within the discourse.

A verbatim transcription was made by the first author of each of the eight retold versions. These versions were analysed for discourse structure, using the method described above. For a more elaborate discussion, see Van Donzel and Koopmans-Van Beinum, 1995b).

Hypotheses

At some points in the discourse, boundaries are more likely to occur than at other points. On the basis of the discourse structure, we expect boundaries to be acoustically realized by the speakers. The question then is: how does the listener perceive these boundaries, on the basis of both semantic and acoustic cues?

We expect boundaries to be perceived at the major transition points in the discourse: phrases, clauses, and paragraphs. At these points, the speaker may signal non-finality (by means of pauses and/or boundary-marking pitch movements), to plan the continuation of his/her retelling, or to give the listener time to process the relation between the preceding clause or phrase and the one he/she is about to utter. The speaker may also signal finality (by means of pauses and/or pitch movements) to indicate the end of a paragraph.

Clause boundaries will then be perceived as non-final. Consequently, boundaries between phrases or after discourse markers and/or connectives will also be perceived as non-final. Boundaries between paragraphs will be perceived as final, since they are heavier (higher in the hierarchical discourse structure).

Listeners, procedure, and listening experiment

Fifteen students of the Free University of Amsterdam were selected as listeners in the perception experiment. They were all familiar with written discourse structures; they all followed an introductory course in discourse analysis. However, they were not discourse experts or trained listeners.
The material consisted of the eight retold versions of the same story, as described above. In total, this was about 30 minutes of speech, including a version used for practice. The eight retold versions were put on audio tape, and sent to the listeners, with detailed information and instructions to the task. This take-home procedure was chosen to avoid too long listening sessions, and to enable the listeners to work on the task in several shorter periods.

The listeners were asked to evaluate the retold versions in terms of phrasing, using only the speech signal. They were asked to assign structure to the text, using conventional punctuation, on the assumption that in a written text, commas correspond to non-final boundaries, and periods correspond to final boundaries. To mark paragraphs, a double slash was used. The relation between punctuation and the discourse boundaries was not mentioned explicitly.

The verbatim transcription of each text without punctuation and capitals, was used as an answer sheet. The listeners were paid for their participation, but only after a careful check by the first author as to the accuracy of the evaluation (in following the instructions). This resulted in the elimination of three evaluations, leaving twelve valid ones.

Results

Each text was evaluated by all 12 listeners. For the processing of the results, the prosody-independent discourse analysis of each of the eight retold versions was taken as a reference point. The perceptual judgements of each listener were marked in the analysis, specifying the exact place and type of each boundary. This resulted in one analysis per version, in which it became clear how many listeners marked a certain point in the discourse with a certain kind of boundary.

Perceived boundaries and discourse structure

First of all, we want to know at what points in the discourse listeners perceive a boundary. At certain points in the discourse, we would expect boundary perception to be more likely than at other points, as explained above. To find out whether our data meet the hypotheses formulated above, we checked for each perceived boundary in each text the number of times it was marked as non-final or final, relative to the discourse structure. Since these types of structures are the same for all speakers, we will take all eight speakers together. This is shown in Table 1.

<table>
<thead>
<tr>
<th>discourse structure</th>
<th>non-final</th>
<th>final</th>
<th>utt. final</th>
<th>par.</th>
</tr>
</thead>
<tbody>
<tr>
<td>dm/conn</td>
<td>284(9)</td>
<td>2(-)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>phrase</td>
<td>876(29)</td>
<td>35(2)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>clause</td>
<td>1908(62)</td>
<td>726(48)</td>
<td>41(27)</td>
<td></td>
</tr>
<tr>
<td>paragr.</td>
<td>751(50)</td>
<td>111(73)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>total</td>
<td>3068</td>
<td>1514</td>
<td>152</td>
<td></td>
</tr>
</tbody>
</table>

The data clearly show that 'non-finality' is mostly perceived at clause boundaries, less at phrase boundaries, and even less at boundaries between discourse markers (dm) and connectives (conn). Non-finality is not perceived at paragraph boundaries. As for 'utterance finality' (marked with a period by the listeners), this type of boundary is equally perceived at clause and at paragraph boundaries. Only few boundaries of finality were perceived after discourse markers or between phrases. Finality associated with paragraphs is perceived mostly at discourse paragraph boundaries.

The results are thus as expected, and meet our hypotheses fairly well. However, some remarks are in order here. The data as presented here do not give any insight into the perceived boundaries for the eight speakers separately. A closer look at the individual speakers revealed that boundary perception is speaker dependent: for some speakers, listeners perceive almost no boundaries at phrase level or after discourse markers, whereas other speakers seem to place their boundaries more at the phrase level than at clause or paragraph level.

Agreement among listeners

A second question we would like to answer is whether the 12 listeners agree in any way on the different types of boundaries. Therefore, we assigned to each perceived boundary a value, relative to the number of responses and the type of boundary marked by the listeners. The assumption is that, the more listeners agree on a specific boundary, the higher the value of that boundary, and the more important it is to the discourse. Boundaries perceived as non-final received 1 point per judgement, those perceived as clause-final received 2 points. Boundaries perceived as
paragraph-final received an additional 3 points per judgement.

As a consequence, the maximum value for a non-final boundary is 12, and for a clause-final boundary 24. Only eight out of 12 listeners assigned paragraph boundaries at all, so the maximum value for a paragraph final boundary is 48. We predict agreement as follows: 1-12 for phrase boundaries, 13-24 for clause boundaries, and 25-48 for paragraph boundaries. Table 2 presents the data, for all speakers and listeners together.

Table 2. Number of perceived boundaries within each category of discourse boundaries, broken down for different categories of agreement, for all speakers and all listeners.

<table>
<thead>
<tr>
<th>category</th>
<th>phrase</th>
<th>clause</th>
<th>paragr</th>
</tr>
</thead>
<tbody>
<tr>
<td>non-final</td>
<td>484</td>
<td>319</td>
<td>2</td>
</tr>
<tr>
<td>(12x1)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>clause final</td>
<td>1</td>
<td>93</td>
<td>46</td>
</tr>
<tr>
<td>(12x2)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>par. final</td>
<td>0</td>
<td>9</td>
<td>36</td>
</tr>
<tr>
<td>(12x2 + 8x3)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The data from table 2 clearly show that the listeners agree in the predicted direction. Phrase and clause boundaries receive highest agreement for 'non-finality', whereas paragraph boundaries receive highest agreement for both clause and paragraph finality.

Conclusions and discussion

Not all listeners marked paragraph boundaries in a systematic way. This could mean that paragraphs do not really exist as such in spontaneous speech. There is evidence that the process unit for spontaneous discourse is the clause ('intonation unit', Chafe, 1994). Clauses combined on functional or semantic grounds form an utterance, and this utterance then coincides with units which listeners can process one at a time. The utterance in spontaneous speech then is comparable to the paragraph in written text. The results on the clause and paragraph-final boundaries also point in this direction.

What we can conclude from this experiment, is that listeners are able to perceptually distinguish between different kinds of discourse boundaries in spontaneous speech in Dutch. Moreover, the perceived boundaries coincide fairly well with the ones predicted from the discourse analysis, as the data on agreement show. The results of a similar experiment by Strangert & Heldner (1995) for Swedish, point into the same direction. However, not all discourse structures, as predicted by the analysis, are realized acoustically by all speakers. This could mean that the 'theoretical' discourse structure is overruled by the acoustic signal.

The results of our experiment will be related in more detail to judgements on perceived prominence, and to acoustic data on pausing strategies. Measurements on boundary-marking pitch movements are being carried out at the moment.

Acknowledgements

The data from this experiment were analysed during the first authors stay at the Department of Linguistics and Phonetics, Lund University (Spring 1996). Many thanks are due to Gösta Bruce and Merle Home for careful reading of the draft.

References


