

# Perception of discourse boundaries and prominence in spontaneous Dutch speech<sup>1</sup>

Monique E. van Donzel<sup>2</sup>

This paper describes the results of a perception experiment in which subjects listening to a retold story in Dutch were asked to mark various perceived discourse structures and to give their judgements on prominence in the verbatim transcription of that story. They marked discourse boundaries by means of conventional punctuation, indicating perceived non-final, sentence final, and paragraph final boundaries. Previously, the written versions had been analyzed for discourse structure on independent, non-prosodic grounds by experts in discourse analysis. This was done at a global level (phrasing) as well as at a local level (focal structure).

The aim of the experiment was to see in what way the perceived structure of a discourse, in terms of boundaries and prominence, coincides with the one predicted by the objectively determined discourse structure.

## Introduction

It is generally assumed that in spoken discourse speakers may use various acoustic means to assign structure to the text, for instance by chunking the text into smaller pieces and by marking highly important words as more prominent. 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 discourse, as used in normal, communicative situations, speakers introduce concepts at various points in the discourse. Depending on the kind of information to be expressed in these concepts, a speaker will mark some concepts as more prominent than others. Concepts introducing 'new' information will generally be more prominent, both from the speaker's and the listener's point of view, than concepts expressing information that has been mentioned previously. A concept introduced

---

<sup>1</sup>Parts of this paper were presented at Fonetik-96, 29-31 May 1996, Nässlingen, Sweden, and at the International Conference of Spoken Language Processing, 3-6 October 1996, Philadelphia, USA (Van Donzel & Koopmans-van Beinum 1996a, 1996b).

<sup>2</sup>Visiting researcher from the Institute of Phonetic Sciences, University of Amsterdam, Herengracht 338, 1016 CG Amsterdam, The Netherlands; vandonzel@fon.jet.uva.nl.

earlier can be referred to by the speaker at later points in the discourse, for instance through the use of pronouns and anaphora (the given/new distinction, cf. Prince 1981, 1992). 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 for the listener to do so, as it would not otherwise be possible to process the pronouns in a meaningful way.

At some points in a discourse, however, pronominalization is not an available option for referring 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 is some kind of discourse boundary, across which pronominalization is not possible (cf. also Horne et al. 1993). This indicates that listeners also have to keep track of more global structures.

Listeners are able to tell when new discourse units begin, and when they end. They also are able to tell what information is most important to the discourse, and which information is of lesser importance. This has convincingly been shown, at least for structured discourse tasks in Dutch, by e.g. Swerts 1994 and Blaauw 1995.

#### *Prosodically marked boundaries*

The prosodic organization of texts read aloud has been widely 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 pre-boundary lengthening, variation in intensity, and sloppy versus clear pronunciation. Full prosodic boundaries are said to include both melodic and pausal cues, while less heavy boundaries are marked with only a pitch movement or a pause (Blaauw 1995). 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 texts read aloud, phrases (noun phrases with determiners and modifiers) and clauses (noun phrases grouped together on semantic, syntactic, or functional grounds) are marked as non-final, while sentences and paragraphs (clauses grouped together as dealing with the same topic) are marked as final by the speaker. This enables the listener to keep track of the global structure of the discourse. Sanderman 1996 showed that listeners are indeed able to distinguish

between boundaries of different 'depths'. She mainly concentrates on sentence initial word boundaries. Swerts 1997 shows similar results for unrestricted spontaneous Dutch monologues, using a method which takes the variation between labellers in segmenting a discourse (conditions being written text alone and written text combined with speech) as a reference point.

Spontaneous discourse, even if it is less strictly structured and characteristically contains more disfluencies and hesitations, will have a structure along the same general 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.

#### *Prominence*

In order to draw the listener's attention to some information in the discourse, a speaker may choose to highlight it. In such a case, at least for Dutch, the part to be emphasized is often realized acoustically with a pitch accent, while duration, amplitude, sloppy vs. clear pronunciation, and voice quality also play an important role (e.g. Koopmans-van Beinum 1992, Rump 1996). A listener is urged to infer from these acoustic cues that the speaker is focusing on a specific part of a linguistic unit. The term 'prominence' is used to refer to the perceptual salience of such a unit relative to its neighbours.

Pitch accent in Dutch is characteristically realized by a so-called 'pointed hat' (cf. 't Hart, Collier & Cohen 1990). A pointed hat consists of an acute accent-lending rise of the fundamental frequency early in the syllable, followed by an acute fall within the same syllable, thus creating the pointed hat contour. The combination of the rise and the fall within the same syllable gives the syllable perceptual salience. The accent-lending rise and fall may also be realized on different syllables, with any number of words intervening. This pattern is called a 'flat hat' contour. Earlier studies showed that the flat hat contour is less frequent in spontaneous speech than in read speech, which is, among other aspects, related to the larger amount of look-ahead and pre-planning in read speech (Blaauw 1995). The listener will be able to detect the important information in the discourse, largely based on the speaker's acoustic realization of pitch accents.

A lot of the research described above has been conducted with spontaneous, albeit fairly structured material, such as instruction monologues. In this paper, we will address the question of how prominence and discourse boundaries are perceived in *unrestricted* spontaneous Dutch speech, as opposed to the more structured type of spontaneous speech.

### Goal of the experiment

The goal of the experiment described here was to investigate how discourse structure, and more specifically *discourse boundaries* and *prominence*, are perceived in spontaneous speech in Dutch. Listeners have to infer the intended global and local structure from the acoustic aspects, on the assumption that speakers mark the different chunks and the important parts of the discourse acoustically, for instance by using pauses. The listener needs this information to process the discourse in a coherent and meaningful way, and to know where to expect new discourse units and salient information. The *perceived* structure indicated by the listeners in the experiment will be related to the objective prosody-independent structure as obtained from the discourse analysis by the experts. We will focus on both global and local discourse features.

### Methods

#### *Speakers, material, and recordings*

Eight native speakers of standard 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 1966). 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 narrative situation. All recordings were made in a sound-treated room on DAT-tape, using a Sennheiser MKH105 high frequency condensator microphone. The retold versions were stored as digitized audio files (sample rate 48 kHz, 16-bit precision).

#### *Discourse analysis*

An independent framework for discourse analysis has been developed previously (Van Donzel 1994, Van Donzel & 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. This method of analysis is inspired by the work of, for instance, Mann & Thompson 1988, Prince 1981, 1992, Chafe 1987, 1994, and Redeker, to appear. First of all, a division is made on a global level. A *clause* is defined as a unit containing words or word groups, which are grouped together on semantic, syntactic, or functional grounds. A *sentence* may contain several clauses, which form a functional unity together. A *paragraph* consists of one or more sentences dealing with the same topic. At the local level, a distinction is made between three types of information: *new information*, i.e. information which the speaker assumes to be completely new in the listeners' context (discourse-new and hearer-new); *inferable information*, i.e. information which the speaker assumes the listener can infer from the preceding context or his/her knowledge of the world (discourse-new, but hearer-old); *evoked information*, i.e. information that has previously been mentioned in the discourse, and that is known to the listener (discourse-old and hearer-old). Furthermore, *discourse markers* (for instance 'but', 'then', 'so', 'anyway'), indicating the main transition points between the different parts of the discourse, were indicated. A distinction was made between 'structural discourse markers', which express some kind of temporal relation between the preceding and following clause, such as 'but', 'afterwards', 'however', and 'pragmatic discourse markers', such as 'well' and 'so', which are more functional in the sense that they reflect information added by the speaker, which is not of crucial importance to the temporal and/or structural organisation of the discourse (e.g. Redeker, to appear). Elements which indicate degrees of manner or place are labelled as *modifiers*. These include mainly adverbs which give additional information about a noun phrase or a verb phrase. *Verbs* were labelled separately; where a verb particle was separated from the verb itself, it was labelled separately as well. The application of this method results in the division of the discourse in terms of clauses, sentences, and paragraphs on a global level, and on a local level in terms of various types of information structure.

A verbatim transcription was made by the author of each of the eight retold versions. These versions were analyzed for discourse structure by the author, using the method described above. In a second stage the analysis was discussed with four discourse experts, and where necessary the analysis was adapted. For a more elaborate discussion, see Van Donzel & Koopmans-Van Beinum 1995b.

## Hypotheses

Boundaries are more likely to occur at some points in the discourse than at other points. We expect the speaker to have realized some of the boundaries acoustically on the basis of the discourse structure to be transferred to the listener. The question then is: how does the listener *perceive* these boundaries, on the basis of both semantic and acoustic cues? Also, some concepts are more likely to be produced as prominent by the speaker than others. Concepts expressing new information according to the discourse analysis are expected to be realized more prominently than the ones referring to given or inferrable information. The question here then is: where does the listener perceive *prominence*, on the basis of both semantic and acoustic cues?

We expect boundaries to be perceived at the major transition points in the discourse: between noun or verb phrases, clauses, sentences, and paragraphs. At these transition points, the speaker may signal *non-finality* or *finality*. A speaker may signal non-finality by means of pauses and/or boundary-marking pitch movements, to plan or to indicate the intention of 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 or discourse unit.

Boundaries after discourse markers and/or connectives, or between phrases will then be perceived as non-final. A clause boundary will be perceived as non-final if it is not the last one in a sentence or a paragraph. Otherwise it will be perceived as final: *utterance final* if it is the last one in a sentence, and *discourse unit final* if it closes a paragraph. Boundaries between paragraphs will also be perceived as final because they are heavier (higher in the hierarchical discourse structure). This is shown graphically in table 1. The acoustic realization of these structures will be investigated at a later stage, and is thus not included in this table.

As for prominence, we expect new information to be perceived as prominent more often than given information, since the former will be marked by the speaker as being of more importance to the listener than the latter. New information will have a 'high' degree of prominence, whereas evoked information will have a 'low' degree. Inferrable information, which the speaker assumes that the listener can infer from the preceding context and his/her knowledge of the world, will then not be marked by the speaker

Table 1. Hypothesized relationship between discourse structure and perceived boundaries.

Discourse structure	Non-final	Utterance final	Discourse unit final
Discourse markers	x		
Noun or verb phrase	x		
Clause	x	x	x
Sentence		x	x
Paragraph			x

as prominently as new information. Since this type of information is still important to the whole discourse, we expect it to be prominent less often than new information, but more often than given information. Thus, inferrable information will have a 'mid' degree of prominence. Discourse markers, which indicate the major transition points between the different parts of the discourse will be marked by the speaker as important, and thus have a 'high' degree of prominence. Modifiers give additional information about a type of information which is not of crucial importance, and can thus be expected to have a 'mid' degree of prominence. This hypothetical relationship is shown graphically in table 2.

## Listening experiment

### *Listeners, procedure, and listening experiment*

Thirteen students of the Free University of Amsterdam and three students of the University of Amsterdam were selected as listeners in the perception experiment. They were all familiar with theories of written discourse structures; they all followed an introductory course in discourse analysis. However, they were not discourse experts or phonetically 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 regarding the task. This take-home procedure was chosen to avoid long listening sessions, and to enable the listeners to work on the task in several shorter periods as accurately as possible.

The listeners were asked to evaluate the retold versions in terms of phrasing, on the basis of 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

**Table 2.** Hypothesized relationship between information structure and perceived prominence.

Information type	Degree of prominence		
	High	Mid	Low
New	x		
Inferrable		x	
Evoked			x
Discourse marker	x		
Modifier		x	

correspond to final boundaries. To mark paragraphs, a double slash had to be used. The relationship between punctuation and discourse boundaries was not mentioned explicitly. Furthermore, the listeners were asked to mark prominence by underlining those words or word groups they perceived as being emphasized by the speaker.

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

## Results on perceived boundaries

### *Listeners' judgements*

Each text was evaluated by all twelve listeners. For the processing of the results, the prosody-independent text-based discourse analysis of each of the eight retold versions was taken as a reference point. The perceptual judgements of the listeners were marked, specifying the exact place and type of each boundary, and the exact words marked as prominent. This resulted in one overview per speaker, in which it became clear how many places in the discourse were marked with a certain kind of boundary by one or more listeners (perceived boundaries), and which words or groups of words were perceived as prominent by one or more listeners (perceived prominence). The results thus obtained were used for further examination.

### *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

**Table 3.** Distribution of total number of places of perceived (perc.) and structural (struct.) discourse boundaries for each type of boundary, broken down for speaker.

Speaker	Non-final		Utterance final		Discourse unit final	
	Perc.	Struct.	Perc.	Struct.	Perc.	Struct.
1	123	43	38	17	12	13
2	110	40	44	20	9	11
3	134	44	45	21	8	11
4	137	41	42	19	9	10
5	116	38	40	15	6	11
6	88	24	29	12	10	10
7	108	28	31	24	6	9
8	141	37	30	18	3	12
total	957	295	299	146	63	87

above. Table 3 presents all places of perceived and structural discourse boundaries, for the eight speakers separately, and broken down for type of perceived boundary. Thus, a 'non-final' boundary is perceived by at least one listener as such in 123 places in the discourse for speaker 1, in 110 places in the discourse for speaker 2, etc. The data are thus *not* weighted for the number of listeners that indicated a specific boundary.

The data from this table show that the perceived number of boundaries for the categories 'non-final' and 'utterance final' exceed the total number of structural boundaries for all speakers. This indicates that listeners also marked these types of boundaries at places other than the actual boundary as predicted by the discourse analysis, i.e. clause-internally. Especially for the 'non-final' boundaries this is not surprising, since non-final boundaries can of course also occur at the phrase level, for instance between noun phrases and/or after discourse markers.

For the 'utterance final' boundaries, it might have been the case that, for instance, the presence of a long pause made the listener mark a specific boundary as utterance final which would not be marked as such in the structural discourse analysis. The exact influence of pausing on the perception of discourse boundaries is still under investigation. The data on 'discourse unit final' boundaries are much clearer: not all structural discourse boundaries, as predicted by the analysis, are perceived as such. Listeners never perceived discourse unit final boundaries at other places in the discourse than the predicted ones.

**Table 4.** Distribution of perceived discourse boundaries relative to structural discourse boundaries, for those cases in which the perceived boundary coincides with a structural one, broken down for speaker.

Struct. Per- ceived Speaker	Clause			Sentence			Discourse unit		
	Non- final	Utt. final	Par. final	Non- final	Utt. final	Par. final	Non- final	Utt. final	Par. final
1	38	9	0	14	13	2	8	13	10
2	37	14	0	17	17	5	8	11	4
3	36	13	0	20	18	2	7	11	5
4	39	13	0	18	16	3	4	9	6
5	36	14	1	15	14	1	10	11	4
6	23	6	0	11	11	3	7	10	7
7	24	2	0	22	19	1	8	8	5
8	36	3	0	18	10	0	5	12	3
total	269	74	1	135	118	17	57	85	44

A chi-square test revealed that there is no effect of speaker on the total number of places where boundaries have been perceived ( $\chi^2=15.30$ ,  $df=14$ ,  $p=.3575$ ). This means that the listeners evaluated all speakers in the same way. For all speakers, listeners perceived roughly three times more places 'non-final' boundaries than where structural boundaries were predicted by the analysis. For all speakers, utterance final judgements are about a factor of two higher than the structural ones predicted. Speakers did not differ significantly in the total number of structural boundaries ( $\chi^2=7.10$ ,  $df=14$ ,  $p=.9305$ ), meaning that the different discourses are comparable in a sensible way.

As can be seen in table 3, many more boundaries were perceived than could be expected on the basis of the discourse analysis. For those places where the perceived boundary coincides with a structural one, we are interested in finding out if the perceived boundary is of the same type as the structural one. In order to investigate this, we selected the places in the discourse where a perceived and a structural boundary coincided. For each type of perceived boundary (non-final, utterance final, and discourse unit final; mutually exclusive), the corresponding structural boundary was checked. This is shown in table 4. The numbers indicate that at least one listener marked a certain boundary at a particular (structural) location in the discourse. Again, the data are not weighted for number of listeners.

The data from this table clearly show that clause boundaries are mostly associated with 'non-finality', and hardly ever with 'paragraph finality'. For

instance, there are 38 places in the discourse where at least one listener has marked a 'non-final' boundary. Occasionally, clause boundaries are perceived as 'utterance final'. For the discourse structure 'sentence', the situation is less clear: listeners tend to perceive 'non-finality' as well as 'utterance finality' at this place. Sometimes, 'paragraph finality' is perceived. At the end of discourse units, listeners perceive mostly 'utterance finality' and 'non-finality'. 'Paragraph finality' is marked in only relatively few cases. The results are thus partly as expected, and agree with our hypotheses fairly well, except for the discourse unit final structure.

An additional remark is appropriate here concerning the presentation of the data. In theory, the total number of perceived boundaries for the three types of discourse boundaries in table 4 should add up to the total number of perceived boundaries in table 3. However, if some listeners marked a specific boundary as 'non-final' and others marked it as 'utterance final', that specific boundary was counted as 'non-final' *and* as 'utterance final'. Statistical analyses will yield an interpretable outcome only when the number of listeners is taken into account, and are therefore not included here. These analyses will of course be included in future experiments.

#### *Agreement among listeners*

A second question we would like to answer is whether the twelve listeners agree in any way on the different types of boundaries. The total number of listeners per judgement in relation to the percentage of judgements explained by this number can be helpful in this respect. We expect much more variation in the number of 'light' boundaries (fewer listeners per boundary) than in the number of 'heavy' boundaries (more listeners per boundary; cf. also Swerts 1997). In other words, we expect that the majority of the judgements consist of ones with lower values for 'non-final' boundaries, and with higher values for 'final' boundaries.

Inspection of the data revealed that there is large variation in the number of listeners per judgement. As a reference point we take the judgements of eight listeners. A frequency count showed that, for the 'non-final' judgements, eight listeners accounted for 95% of all judgements. For the 'utterance final' judgements, they accounted for 72%, and for the 'paragraph final' judgements they accounted for 97%. These figures indicate that agreement among listeners is rather high, especially for the 'non-final' and 'paragraph final' judgements. It is less for 'utterance final' judgements, but still reasonably high.

The results presented so far concerned the perception of discourse structure at a global level: phrasing. We will now turn to the results of perceived structure at a local level: prominence.

## Results for perceived prominence

### *Listeners' judgements*

For the processing of the data on perceived prominence, the same overview of perception judgements per speaker was used as for the boundaries. This overview also contained information on how many listeners perceived a certain word or group of words as prominent. For further details, see the section on boundaries.

### *Perceived prominence and information structure*

To analyze the scores in perceived prominence, we took the overall analyses of the eight versions as a reference point. The maximum score for each item perceived as prominent was twelve, since twelve listeners participated in the experiment. Obviously, not all listeners marked an equal number of items in the discourse as prominent. Some listeners evaluated the discourses in much more detail than others. Since we are mainly interested in the items perceived as prominent by the majority of listeners, we selected those items which received at least eight judgements of prominence (out of the total of twelve). These items then represent the ones on which listeners agreed most.

For each of these perceptually prominent items, we checked the information status. Table 5 shows the total number of words marked as prominent by one or more listeners, as well as the selection of prominent words which received at least eight judgements, for the eight speakers separately.

These figures make sense only when we have information about the total number of items within each category, based on the prosody-independent discourse analysis. This is shown in table 6 for the essential categories of information structure. Total numbers of evoked items are included in the rest category, since evoked information includes all pronouns, and since the percentage of prominent pronouns is negligible. The percentage perceived as prominent ( $\geq 8$  judgements) within each category is included. The total number of prominent items does not differ significantly for the eight speakers ( $\chi^2=36.95$ ,  $df=28$ ,  $p=.12$ ).

Note that the total number of perceived words (table 5) may exceed the total number of items per category (table 6), since an item or concept may

**Table 5.** Distribution of total number of words perceived as prominent ( $\geq 1$  judgement), and of number of words perceived as 'highly' prominent ( $\geq 8$  judgements), broken down for information type and speaker. The information types are coded as follows: new = new information, inf = inferrable information, dm = structural and pragmatic discourse markers, mod = modifier, verbs = auxiliaries, main verbs and verb particles; the rest group includes evoked information, determiners, and *wh*-words.

Information		Speaker							
		1	2	3	4	5	6	7	8
New	total	39	38	50	43	32	35	32	58
	$\geq 8$	18	10	22	14	9	18	16	18
Inf	total	43	30	39	29	35	33	28	22
	$\geq 8$	10	10	12	10	5	9	7	6
Dm	total	18	8	11	13	10	5	9	6
	$\geq 8$	3	1	2	0	0	0	1	0
Mod	total	30	21	40	26	31	13	25	25
	$\geq 8$	5	1	15	2	4	0	2	2
Verb	total	54	41	40	43	30	26	37	41
	$\geq 8$	6	5	4	10	4	3	2	8
Rest	total	11	5	5	8	12	2	8	11
	$\geq 8$	0	0	2	0	0	0	1	2
Sum	total	195	143	185	162	150	114	139	163
	$\geq 8$	42	27	57	36	22	30	30	35

consist of several words. In the experts' analysis, concepts are labelled, rather than words. In the prominence analysis, the words within one concept were counted separately if the judgements for prominence differed. Concepts can be compared to noun phrases (in syntactic terms): a noun accompanied by determiners, modifiers, and/or adjectives.

Table 6 shows the total number of concepts based on the discourse analysis, and the percentage perceived as prominent ( $\geq 8$  judgements) per information category. A chi-square test revealed that the total number of concepts does not differ significantly for the eight speakers ( $\chi^2=27.07$ ,  $df=28$ ,  $p=.514$ ). The mean percentage of prominence in each information category indicates that 48% of all new items are perceived as prominent, 26% of all inferrable items, 3% of all discourse markers, and 7% of all modifiers. The speaker differences are substantial. If we look at the 'new' information, for example, we see 69% prominent for speaker 4, compared to 26% for speaker 2. The category 'modifiers' is even clearer: 22% for speaker 3 versus none for speaker 6. These differences could very well be a

**Table 6.** Total absolute number of concepts based on the discourse analysis, and percentage perceived as prominent ( $\geq 8$  judgements) in each information category, broken down for speaker. The information types are coded as in table 5.

Information	Speaker								mean	
	1	2	3	4	5	6	7	8		
New	total	37	38	39	32	31	30	27	46	35
	%prom	49	26	56	69	29	60	59	39	48
Inf	total	35	32	44	33	34	32	25	28	33
	%prom	28	31	27	30	15	28	28	21	26
Dm	total	35	24	36	44	48	19	29	47	35
	%prom	8	4	6	0	0	0	3	0	3
Mod	total	48	50	69	51	59	28	50	61	52
	%prom	10	2	22	4	7	0	4	3	7
Verb	total	77	70	81	75	76	48	76	75	72
	%prom	8	7	5	13	5	6	3	11	7

matter of 'good' or 'bad' retellers. In order to find out the listeners' preference for the speaker-specific ways of retelling, we performed a listening experiment in which listeners were asked to evaluate the retelling task of each speaker using different semantic scales expressing prosodic characteristics. Preliminary results indicate that the number of modifiers used by the speaker is representative of a 'good' or 'lively' speaking style (speaker 3 receives the highest overall score; speaker 6 is second-last).

## Conclusions and discussion

### *Phrasing*

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 processing unit for spontaneous discourse is the clause ('intonation unit', Chafe 1994). Clauses combined on functional or semantic grounds form a sentence or discourse unit, and this unit then coincides with units which listeners can process one at a time. The sentence in spontaneous speech is thus comparable to the paragraph in written text. The results at the clause- and paragraph-final boundaries also point in this direction.

What we can conclude from the data on boundary perception is that listeners are able to perceptually distinguish 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 in the same direction. However, not all discourse structures, as predicted by the analysis, need to be realized acoustically by all speakers. This could mean that the 'theoretical' discourse structure is overruled by the 'actual' acoustic signal.

### *Prominence*

The data on prominence show a number of interesting observations. First of all, if we compare the total number of items perceived as prominent to the number of items where at least eight listeners agree, we see that the majority of the judgements are not captured. This means that there is variation among the listeners as to which items are prominent. The category ' $\geq 8$ ' ('high') reflects those items on which listeners agree most. The category '<8' ('mid') will then have to be examined in much more detail to find out why listeners marked these items as prominent in a less homogeneous way.

Another point is the speaker differences at information-type level. Although the total distribution of items per information category is not significantly different for the eight speakers, the number of discourse markers and modifiers differs substantially from one speaker to the next, ranging from 19 to 48 and from 28 to 69 respectively (see table 6). These elements are used by the speaker to give the discourse a 'personal touch'. The 'basic' elements which build up the discourse, such as new and inferrable information, show fewer differences between the speakers. Of course, all speakers have realized the discourse in their own way, and differences in length are natural. The speaker differences could very well reflect the speakers' ability to tell a story in a good or bad way. The preliminary results from the evaluation experiment indicate that this seems to be the case.

The percentages perceived as prominent deviate from other findings in the literature (e.g. Brown 1983), which are generally much higher. This might be an effect of speech material: whereas most studies use fairly structured speech, for instance instruction monologues or dialogues, our material is 'more' spontaneous in the sense that it is unrestricted. Language may also be of influence, as most studies have been conducted on spontaneous English production, whereas we used Dutch speech.



**Table 6.** Total absolute number of concepts based on the discourse analysis, and percentage perceived as prominent ( $\geq 8$  judgements) in each information category, broken down for speaker. The information types are coded as in table 5.

Information	Speaker								mean	
	1	2	3	4	5	6	7	8		
New	total	37	38	39	32	31	30	27	46	35
	%prom	49	26	56	69	29	60	59	39	48
Inf	total	35	32	44	33	34	32	25	28	33
	%prom	28	31	27	30	15	28	28	21	26
Dm	total	35	24	36	44	48	19	29	47	35
	%prom	8	4	6	0	0	0	3	0	3
Mod	total	48	50	69	51	59	28	50	61	52
	%prom	10	2	22	4	7	0	4	3	7
Verb	total	77	70	81	75	76	48	76	75	72
	%prom	8	7	5	13	5	6	3	11	7

matter of 'good' or 'bad' retellers. In order to find out the listeners' preference for the speaker-specific ways of retelling, we performed a listening experiment in which listeners were asked to evaluate the retelling task of each speaker using different semantic scales expressing prosodic characteristics. Preliminary results indicate that the number of modifiers used by the speaker is representative of a 'good' or 'lively' speaking style (speaker 3 receives the highest overall score; speaker 6 is second-last).

## Conclusions and discussion

### *Phrasing*

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 processing unit for spontaneous discourse is the clause ('intonation unit', Chafe 1994). Clauses combined on functional or semantic grounds form a sentence or discourse unit, and this unit then coincides with units which listeners can process one at a time. The sentence in spontaneous speech is thus comparable to the paragraph in written text. The results at the clause- and paragraph-final boundaries also point in this direction.

What we can conclude from the data on boundary perception is that listeners are able to perceptually distinguish 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 in the same direction. However, not all discourse structures, as predicted by the analysis, need to be realized acoustically by all speakers. This could mean that the 'theoretical' discourse structure is overruled by the 'actual' acoustic signal.

### *Prominence*

The data on prominence show a number of interesting observations. First of all, if we compare the total number of items perceived as prominent to the number of items where at least eight listeners agree, we see that the majority of the judgements are not captured. This means that there is variation among the listeners as to which items are prominent. The category ' $\geq 8$ ' ('high') reflects those items on which listeners agree most. The category '<8' ('mid') will then have to be examined in much more detail to find out why listeners marked these items as prominent in a less homogeneous way.

Another point is the speaker differences at information-type level. Although the total distribution of items per information category is not significantly different for the eight speakers, the number of discourse markers and modifiers differs substantially from one speaker to the next, ranging from 19 to 48 and from 28 to 69 respectively (see table 6). These elements are used by the speaker to give the discourse a 'personal touch'. The 'basic' elements which build up the discourse, such as new and inferrable information, show fewer differences between the speakers. Of course, all speakers have realized the discourse in their own way, and differences in length are natural. The speaker differences could very well reflect the speakers' ability to tell a story in a good or bad way. The preliminary results from the evaluation experiment indicate that this seems to be the case.

The percentages perceived as prominent deviate from other findings in the literature (e.g. Brown 1983), which are generally much higher. This might be an effect of speech material: whereas most studies use fairly structured speech, for instance instruction monologues or dialogues, our material is 'more' spontaneous in the sense that it is unrestricted. Language may also be of influence, as most studies have been conducted on spontaneous English production, whereas we used Dutch speech.

### Implications and future research

In this paper we presented the results of a perception experiment which was carried out to investigate the relationship between the perceived and the objectively determined discourse structure. We focused on only the perceptual judgements. The acoustic part of the issue has not been addressed, since these data are currently being processed. One of the questions to be answered in the near future, then, is what acoustic cue made the listeners mark an item as prominent. A first look at the F0 data suggests that nearly all prominent items were realized with a pitch accent. These findings are in agreement with the results from experiments performed on the read versions of the material used in the present experiment (Streefkerk & Pols 1996). The number of prominences realized without pitch accent is negligible. This does not mean, however, that listeners scored pitch movements only, nor that pitch is the only acoustic determinant of prominence.

The experiment described in this paper is part of a larger project on the acoustic-phonetic determinants of focusing in discourse. Other experiments performed within this project include the evaluation of prosodic characteristics in retold stories using semantic scales (Van Donzel & Koopmans-van Beinum 1997a), the perceptual difference between spontaneous and read speech (Van Donzel & Koopmans-van Beinum 1995b, Koopmans-van Beinum & Van Donzel 1996), and acoustic measures of pause duration (Van Donzel & Koopmans-van Beinum 1996b), speaking rate (Koopmans-van Beinum & Van Donzel 1996), and fundamental frequency (Van Donzel & Koopmans-van Beinum 1997b). The contribution of these measures will be included at a later stage, in order to provide a complete picture of the acoustic realization and perceived structure of discourse in Dutch. Additional measurements and analyses of the F0-data are being carried out at the moment.

### Acknowledgements

The author's stay at the Department of Linguistics and Phonetics, Lund University (Spring 1996), was financed by the Netherlands Organisation for Scientific Research (NWO), and the ERASMUS Network 'Phonetics and Speech Communication'. Many thanks are due to Gösta Bruce, Merle Horne, and Florian Koopmans-van Beinum for careful reading of and comments on earlier versions of parts of this paper, and to the whole Department of Linguistics and Phonetics, Lund University, for fruitful discussions, and a really great time.

### References

- Blaauw, E. 1995. *On the perceptual classification of read and spontaneous speech*. Doctoral dissertation, Utrecht University.
- Brown, G. 1983. 'Prosodic structure and the given/new distinction'. In A. Cutler & D.R. Ladd (eds.), *Prosody: Models and measurements (Springer Series in Language and Communication 14)*, 67-77. New York: Springer-Verlag.
- Bruce, G. 1982. 'Textual aspects of prosody in Swedish'. *Phonetica* 39, 274-87.
- Carmiggelt, S. 1966. 'Een triomf'. *Fluiten in het donker*. Amsterdam: ABC Boeken.
- Chafe, W.L. 1987. 'Cognitive constraints on information flow'. In R.S. Tomlin (ed.), *Coherence and grounding in discourse (Typological studies in language 11)*, 21-51. Amsterdam: John Benjamins.
- Chafe, W.L. 1994. *Discourse, consciousness, and time*. Chicago: The University of Chicago Press.
- 't Hart, J. R. Collier & A. Cohen. 1990. *A perceptual study of intonation. An experimental-phonetic approach to speech melody*. Cambridge: Cambridge University Press.
- Horne, M., M. Filipsson, M. Ljungqvist & A. Lindström. 1993. 'Referent tracking in restricted texts using a lemmatized lexicon: Implications for generation of intonation'. *Proceedings Eurospeech '93*, 2011-14. Berlin.
- Koopmans-van Beinum, F.J. 1992. 'The role of focus words in natural and in synthetic continuous speech: acoustic aspects'. *Speech Communication* 11, 439-52.
- Koopmans-van Beinum, F.J. & M.E. van Donzel. 1996. 'Discourse structure and its influence on local speech rate'. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam* 20, 1-11.
- Lehiste, I. 1979. 'Perception of sentence and paragraph boundaries'. In B. Lindblom & S. Öhman (eds.), *Frontiers of speech communication research*, 191-201. London: Academic Press.
- Mann, W.C. & S.A. Thompson. 1988. 'Rhetorical structure theory: Toward a functional theory of text organization'. *Text* 8:3, 243-81.
- Nakatani, C.H., J. Hirschberg & B.J. Grosz. 1995. 'Discourse structure in spoken language: Studies on speech corpora'. In *AAAI Spring symposium on empirical methods in discourse interpretation and generation*. Stanford.

- Prince, E. 1981. 'Toward a taxonomy of Given-New information'. In P. Cole (ed.), *Radical Pragmatics*, 223-55. New York: Academic Press.
- Prince, E. 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*, 295-325. Amsterdam: John Benjamins.
- Redeker, G. to appear. 'Coherence and structure in text and discourse'. In W. Black & H. Bunt (eds.), *Computational pragmatics: abduction, belief, and context*. London: University College Press.
- Rump, H.H. 1996. *Prominence of pitch-accented syllables*. Doctoral dissertation, Eindhoven University.
- Sanderman, A.A. 1996. *Prosodic phrasing. Production, perception, acceptability, and comprehension*. Doctoral dissertation, Eindhoven University.
- Strangert, E. & M. Heldner. 1995. 'Labelling of boundaries and prominences by phonetically experienced and non-experienced transcribers'. *PHONUM* 3, 85-109. Umeå University.
- Streefkerk, B.M. & L.C.W. Pols. 1996. 'Prominent accent and pitch movements'. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam* 20, 111-19.
- Swerts, M. 1994. *Prosodic features of discourse units*. Doctoral dissertation, Eindhoven University.
- Swerts, M. 1997. 'Prosodic features at discourse boundaries of different strength'. *Journal of the Acoustical Society of America* 101, 514-521.
- Van Donzel, M.E. 1994. 'How to specify focus without using acoustic features'. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam* 18, 1-17.
- Van Donzel, M.E. & F.J. Koopmans-van Beinum. 1995a. 'Evaluation of discourse structure on the basis of written vs. spoken material'. In K. Elenius & P. Branderud (eds.), *Proceedings XIIIth ICPHs*, 258-61. Stockholm.
- Van Donzel, M.E. & F.J. Koopmans-van Beinum. 1995b. 'Prominence judgements and textual structure in discourse'. *Proceedings of the Institute of Phonetic Sciences, University of Amsterdam* 19, 11-23.
- Van Donzel, M.E. & F.J. Koopmans-van Beinum. 1996a. 'Perception of discourse boundaries in spontaneous speech in Dutch'. *Proceedings Fonetik-96 (TMH-QPSR 1996/2)*, 105-108. KTH, Stockholm.

- Van Donzel, M.E. & F.J. Koopmans-van Beinum. 1996b. 'Pausing strategies in discourse in Dutch'. *Proceedings ICSLP96*, 1029-32. Philadelphia.
- Van Donzel, M.E. & F.J. Koopmans-van Beinum. 1997a. 'Evaluation of prosodic characteristics in retold stories in Dutch by means of semantic scales'. In *Eurospeech '97 Proceedings* vol. 1, 211-14. University of Patras.
- Van Donzel, M.E. & F.J. Koopmans-van Beinum. 1997b. 'Pitch accents, boundary tones, and information structure in spontaneous discourse in Dutch'. In A. Botinis, G. Kouroupetroglou & G. Carayiannis (eds.), *Intonation: Theory, models and applications. Proceedings of an ESCA workshop*, 313-16. University of Athens.