Prosodic Aspects of Information Structure in Discourse
van Donzel, M.E.

Citation for published version (APA):
2 The relation between textual information structure and perceived prominence in discourse

Abstract

This chapter presents an overview of different methods of discourse analysis. Three approaches based on textual representation of discourse are evaluated: Rhetorical Structure Theory, Cognitive constraints of information flow, and Assumed Familiarity.

On the basis of, and inspired by, these approaches a new text-based method is introduced, which is suitable to analyze different features of spontaneous discourse, to be related to prosodic measures afterwards.

This is followed by the presentation of a pilot experiment, in which the proposed method is tested by comparing the text-based discourse segmentation (text only) with a segmentation on the basis of listeners' judgements (text plus prosody).

The results of this experiment prompted us to modify the proposed method on some points. These modifications are discussed, and the current version of the framework is presented. This framework is called Information Structure In Discourse (ISID).

* This chapter is a substantially revised and extended version of Van Donzel (1994) and Van Donzel & Koopmans-van Beinum (1995a and 1995b).
2.1. Introduction

The aim of the research described in this thesis is to investigate the possible ways in which a speaker can mark different levels of 'structure' in spoken discourse, and what cues the listener uses to detect that 'structure'. We will use the term discourse structure to refer to boundary marking, and the term information structure to refer to focus marking.

Focus is usually defined by means of intonation, namely, by stating that a word or a word group is placed in focus if it is realized with a (pitch) accent. Since acoustic parameters are the object of our investigation, however, this kind of definition may lead to circularity: the possible acoustic features are already included in the definition itself. Such circularity can be avoided by looking for a way to define focus without including any intonational or other acoustical features. Therefore, we will need a definition of the notion focus, as well as of other notions related to this, such as old vs. new information, based on the textual structure as such. Chapter 5 deals with the acoustic marking of focal structure.

The same is, of course, true for boundary marking. Prosodic boundaries are said to be acoustically signaled by so-called 'boundary tones' and by pausing. The definition of where discourse boundaries occur can evidently not include terms like boundary tones.

In this chapter, we will concentrate on the approaches to the analysis of the internal focal structure of a text, thus on the (local) word level rather than on the (global) utterance discourse level. The acoustic realization of boundary marking and its relation to discourse structure will be dealt with in more detail in chapter 4.

In writing, in order to draw the reader's attention to any particular information in a written discourse, a writer may use typographical means such as putting important words in boldface or underlining them. If a speaker wants to draw the listener's attention to particular information in spoken discourse, he/she may also choose to highlight it. In such a case, at least for Dutch, the part to be emphasized is often acoustically realized with a pitch accent, while duration, pausing, amplitude, sloppy vs. clear articulation, and voice quality also play an important role (e.g. Koopmans-van Beinum, 1992a, 1992b; 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 in relation to its neighbors.

Basically there are three approaches one can use to analyze the information structure of a spoken sentence or a text with respect to prominence. This is also referred to as the 'accent placement debate' (cf. Baart, 1987):
TEXTUAL INFORMATION STRUCTURE AND PROMINENCE

- **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);
- **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” (ibid., p. 11);
- **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 focused 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 focused constituents are automatically derived on the basis of lexico-syntactic structure” (ibid., p. 11).

Both the syntactic approach 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. The speech material used in such approaches thus consists of the acoustic realization of a written construction the speaker is asked to read aloud. This is not representative of a situation in which a speaker is asked to tell something in his/her own words, i.e. a situation in which the speaker is considering what to say, as well as how to say it. In our view the pragmatic approach is more suitable, yet less specific, to analyze discourse and spontaneous speech, since it takes into account the notions ‘knowledge of the world’ and ‘context’, which are crucial in spontaneous speech.

A specific implementation of an approach to 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, based on implementation rules, automatically break a given text down into different parts conveying different kinds of information, and then assign pitch accents and other prosodic features to the most salient parts. However, at this stage, for texts to be correctly analyzed, the input of the algorithms usually needs to consist of grammatically correctly structured sentences. Thus, this method would not be useful to analyze ‘real’ spontaneous speech, since this type of speech often shows a lot of hesitations and unfinished (‘ungrammatical’) utterances.
These considerations, together with the above-mentioned circularity (the definition of focus already including acoustic features) led us to develop an 'objective' method to analyze discourse and spontaneous speech material, initially independent of acoustic features, by taking into account pragmatic aspects. Subsequently, the material can be analyzed acoustically, in order to investigate the acoustic-phonetic correlates of the objective information structure.

This chapter is organized as follows: Section 2.2 will give an overview of different approaches to discourse structure as found in the literature. Section 2.3 presents a pilot experiment conducted to test the applicability of the approaches found in the literature. Finally, in section 2.4 we present the framework used in this research, allowing us to indicate discourse structure at distinct levels (boundaries and focal structure), independent of prosodic features. The development of this method was inspired by the approaches described in section 2.2. The application of the methods results in an independent analysis of any text, which can easily be compared to the acoustic realization of that text by the speaker on the one hand, and to the perceived structure of that same text by the listener on the other. Figure 2.1 shows the various aspects of this study, and how they relate to each other.

Figure 2.1 Overview of the various aspects in this study, and their relation to one another.
Central to the whole issue is the discourse. 'Discourse' is defined here as a coherent piece of text, either spoken or written, produced by one speaker or writer and consisting of at least one paragraph; it differs from 'dialogue', in which at least two speakers are involved. The speaker makes use of concepts in order to formulate his/her message. The speaker has both linguistic and acoustic means available to structure the concepts into a message, i.e. the spoken discourse. The linguistic and acoustic means are also picked up by the listener listening to the spoken discourse. He/she uses them to reconstruct the message of the speaker (the discourse), this time in terms of percepts (cf. the model of verbal interaction by Dik (1997, part II, p. 410).

The structure of the discourse itself has thus two sides: the structure as produced by the speaker (cf. the 'writer' of the discourse) and that perceived by the listener (cf. the 'reader' of the discourse). Both consist of at least two levels: one to indicate the broad structure of the message in terms of where discourse units begin and end ('discourse structure' in the diagram), and one to indicate the narrow structure in terms of important information ('information structure' in the diagram).

The diagram presented above is crucial to the structure of the whole research. It is focused on the relation between discourse and information structure and its realization by the speaker on the one hand, and its perception by the listener on the other hand.

2.2. Methods of discourse analysis

2.2.1. Introduction

This section presents three representative theories about the structure of written text, based on textual analysis: Rhetorical Structure Theory (RST), introduced by Mann & Thompson (1988), the approach of Chafe (1987), and the one proposed by Prince (1981). This last one is concerned with a smaller domain (words or word groups, in any case NPs) than the former two (whole texts). These three approaches mainly focus on the coherence relations in a discourse, and on the ways to represent them. The starting point for these theories is generally a written text. RST is explicitly designed to analyze texts written with specific goals, such as fund-raising letters. The verbatim transcription of a spoken text is also used as a starting point, as is the case for the theories by Chafe and Prince. Both theories are based on transcribed spontaneous speech. However, no direct relation is given between the features of the text itself and the acoustic realization, while the main focus is on the written version.
RST is introduced as a method to account for the structure of texts ‘primarily in terms of relations that hold between parts of the text’ (Mann & Thompson, 1988, p. 243). This means that RST can be applied to assign structure to a text above the level of the sentence: the text is divided into functional units, and between these units several relations can hold. The result of such an analysis is a very global division in units.

The approach proposed by Chafe (1987) is more specific than RST, in the sense that it focuses on sentences, whereas RST focuses on more global structures such as texts. The theory by Chafe accounts for the status of concepts within so-called ‘intonation units’ (defined auditorily); these concepts can be ‘active’, ‘semi-active’, or ‘inactive’. The method proposed by Prince (1981) is comparable to that of Chafe, but deals with structures in the form of noun phrases. This type of analysis results in a very detailed functional description, and is based on linguistic representation. One fundamental difference between these last two analyses is that Chafe’s analysis concerns the activation state of a referent in the head of the hearer, whereas Prince’s analysis refers to the formulation chosen by the speaker: a referent is classified as ‘brand new’ if the speaker formulates it as ‘brand new’. For an elaborate overview of approaches to text relations and information structure, see Lambrecht (1994), Chafe (1994), and Kroon (1995).

The methods described above are obviously not the only ones available to analyze the structure of discourse. Other methods for the analysis of discourse structure have been proposed by computational linguists, for instance, Polanyi & Schä (1983), Brennan et al. (1987), Passoneau (1996), Vallduvi (1994), Walker et al. (1997), and Grosz & Sidner (1986).

Much work has been done to investigate the relation between discourse structure and the prosodic features, using prosody-independent frameworks of discourse structure (e.g., Geluykens & Swerts, 1994; Grosz & Hirschberg, 1992; Hirschberg & Grosz, 1992, 1994; Nakajima & Allen, 1992; Nakatani, 1995; Nakatani, Hirschberg & Grosz, 1995; Swerts & Geluykens, 1994; Terken & Hirschberg, 1994; Venditti & Swerts, 1996).

In their ‘Instructions for annotating discourse’, Nakatani et al. (1995) present a manual guide for labeling discourse, which can be used by naive subjects. The labeling instructions are based on a specific theory of discourse structure (Grosz & Sidner, 1986), but no explicit reference is made to that theory. The idea is that listeners ‘reason about how the speaker chose to organize the information s/he intended to communicate to the hearer’ (Nakatani et al., p. 1). ‘Listeners’ here refers to the labelers of the discourse structure. Subjects are asked to add indents and labels to a text, in order to mark the structure of the written text. The text has
been divided beforehand into intermediate phrases by those performing the experiment, and the subjects are not allowed to change this division. Thus the application of these instructions only give information about the relation and the hierarchical structure between the different phrases of which the text is built up, that is, the overall global structure of a whole discourse, rather than the information status of the individual words on a more local level. However, Hirschberg & Grosz (1992) analyze their material both on a global level (the structure of the discourse constituents which form the whole discourse) and on a local level (parentheticals, quotations, tags, and indirect reported speech). This definition of 'local' level differs from ours, for by 'local' we mean the structure of different types of information within the utterance.

To account for the more local level of information structure, centering may be used (e.g. Brennan et al., 1987; Passoneau, 1996). Centering can be seen as 'a system of rules and constraints that govern the relationship between what the discourse is about and some of the linguistic choices made by the discourse participants' (Brennan et al., 1987), i.e. the informational content of the discourse. This centering approach has been developed for English; for Dutch no such approach has been developed yet.

In the present thesis, we want to approach the structure of a text from a pragmatic point of view rather than from a computational one. The models and theories presented by Mann & Thompson (1988), Chafe (1987), and Prince (1981) are therefore more appropriate in this respect than the more computationally oriented approaches. However, the former models present either an account of the 'global' structure of a text (i.e. sentences and paragraphs) or the more 'local' structure (i.e. information status). Since we want to analyze both aspects and relate them to the acoustic realizations, we will develop a new independent framework for discourse analysis, to include both levels.

2.2.2. Rhetorical Structure Theory

RST (Mann & Thompson, 1988) is a theory developed to identify the hierarchical structure in a text, to describe relations between text parts and their transitions, and thus to give a comprehensive analysis. It was designed for written monologues. Studies which have used RST revealed a number of advantages: relations among clauses can be described regardless of whether they are grammatically or lexically signaled; RST is applicable to a wide range of text types and to narrative discourse (see, for instance, Mann & Thompson, 1992; Abelen, Redeker & Thompson, 1993; Redeker, to appear).

RST has defined four objects: Relations, Schemas, Schema applications, and Structures.
• **Relations.** Relations hold between two spans of text (non-overlapping) which are called 'nucleus' and 'satellite'. The four fields that each relation consist of are constraints on nucleus, on satellite, on the combination of both, and on the effect. Each field specifies judgements the analyst must make when building the RST structure. These are judgements of plausibility.

• **Schemas.** Schemas refer to constituent arrangements, comparable to grammatical rules. These schemas specify how text spans can co-occur. There are five kinds of schemas. Curved lines represent relations; straight lines represent identification of the nuclear spans. The schemas all follow the pattern of a single relation with a nucleus and a satellite.

• **Schema application.** Schema applications specify the possible applications of a schema: unordered spans (no constraint on the order of nucleus and satellite), optional relations (in multi-relation schemas at least one relation must hold), and repeated relations (a relation can be applied any number of times).

• **Structure.** A text is divided into units, which should have independent functional integrity. These units are usually clauses. The analysis is a set of schema applications which satisfy the following constraints: completeness, connectedness, uniqueness, and adjacency. RST analyses are presented in the form of hierarchical trees.

The definitions used to describe the different relations between clauses are not based on morphological or syntactic signals, but are recognized on the basis of functional and semantic judgements. Relation definitions that can hold between the different parts of a text are for example: Circumstance, Solutionhood, Elaboration, Background, Enablement, Motivation, Evidence, Justification, Relations of cause, Antithesis, Conclusion, Condition, Interpretation, Evaluation, Restatement, Summary, Sequence, and Contrast. This set is considered to be open. Figure 2.2 gives a small example of the relation definition ‘Evidence’, as presented by Mann & Thompson (p. 251); N = nucleus, S = satellite, R = reader, W = writer. The interested reader is referred to Mann & Thompson (1988) for more examples of (larger) text analyses.

The relation definitions described above can be classified in a two-way distinction as subject matter relations and presentational relations. Subject matter relations are defined as “those whose intended effect is that the reader recognizes the relation in question” (p. 257). Presentational relations are “those whose intended effect is to increase some inclination in the reader, such as the
desire to act or the degree of positive regard for, belief in, or acceptance of the nucleus" (p. 257). This division is the one proposed by Mann & Thompson.

<table>
<thead>
<tr>
<th>relation name:</th>
<th>Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>constraints on Nucleus:</td>
<td>R might not believe N to a degree satisfactory to W</td>
</tr>
<tr>
<td>constraints on Satellite:</td>
<td>The reader believes S or will find it credible</td>
</tr>
<tr>
<td>constraints on the N+S combination:</td>
<td>R’s comprehending S increases R’s belief of N</td>
</tr>
<tr>
<td>the effect:</td>
<td>R’s belief of N is increased</td>
</tr>
<tr>
<td>locus of the effect:</td>
<td>N</td>
</tr>
</tbody>
</table>

Figure 2.2 Example of the relation definition 'Evidence', taken from Mann & Thompson (1988, p. 251).

A constraint against inappropriate use of relations is assured by the Effect: “for each relation and schema definition, the definition applies only if it is plausible to the analyst that the writer wanted to use the spanned portion of the text to achieve the Effect” (p. 258). This means that RST structures are structures of functions rather than of forms.

Studies involving the application of RST to natural languages give insight into the use and consequences of RST. Results from text analyses have shown the following (as formulated by Mann & Thompson, p. 259):

- Virtually every text can be given an RST analysis.
- There are certain text types which characteristically cannot be given an RST analysis, for instance laws, contracts, and poetry.\footnote{Mann & Thompson do not treat spoken texts in their corpus. The fact that RST is designed for written monologues implies that some additional features are needed to analyze spontaneously spoken monologues and to characterize all their characteristics, such as hesitations and corrections. Ideally, RST analyses for laws, contracts, and poetry should, of course, also be possible.}
- In our culture, texts having an RST analysis predominate. RST is thus not a universal property of a text.

Results from studies of relational properties show that (pp. 259-260):

- Structural relations are not necessarily expressed in clauses.
- Such propositions may be signaled by conjunctions or other morphemes, but they can also be conveyed without.

1
• The relational propositions correspond to the relations of the RST structures of the text.
• The relational propositions are essential to the coherence of a text: if these are disturbed, the text will become incoherent.

Mann & Thompson also present evidence for nuclearity. The notions of nucleus and satellite were introduced above. The relation between them is not symmetrical: the nucleus is considered to be the central principle around which the text structure is built. This leads to the prediction that if a nucleus is removed, the significance of material in its satellite will not be apparent. The data analyzed by Mann & Thompson show that this prediction is correct: a text consisting of only satellites is incomprehensible and incoherent, and the reader does not have a clear idea what the text is about.

Another prediction is that if the satellite is removed, the text should still be coherent. This prediction is supported as well by the data. These findings present strong evidence for the claim for nuclearity. If communication is seen as ‘building memories’, the function of nuclearity seems to be the organization of details in these memories. The nucleus is the part that is most deserving of response, including attention and reaction. The nucleus is more central than the satellite.

In conclusion we may state that RST turns out to be a very useful method to analyze different types of discourse. It defines the hierarchical structure of texts and describes the relations that hold between the different parts in functional terms. The distinction between nucleus and satellite enables RST to describe clause combining, and thus coherence in discourse.

RST can be applied to analyze a text or discourse on the level above the sentence. This means a rough analysis in terms of ‘functional units’, following the RST rules. These ‘unitization rules’ form a preliminary step, and do not form part of the actual analysis itself. The division of the text above the level of the sentence is needed to account for certain boundary effects. The presence and the place of these boundaries may follow from the RST analysis. Therefore, a rather detailed description of Rhetorical Structure Theory was necessary. Within the parameters of our research, however, it is not crucial to define the relations between the units, since our primary concern is the internal focal structure within clauses or sentences. This is not accounted for by RST, and therefore we will make additional use of the theories of Chafe and Prince to determine the structure of texts on the sentence level and below.
2.2.3. Cognitive constraints of information flow

Chafe (1987) proposes an approach to analyze the information flow in terms of cognitive constraints. Chafe’s terminology may suggest that the analysis is done on the basis of acoustic features. We feel, however, that this theory can be of use in our research, because the basis of the theory is the analysis of a transcribed spontaneously uttered text rather than the acoustic parameters of the speech signal.

A piece of (transcribed) spoken language naturally divides itself into ‘intonation units’ (a single focus of a speaker’s consciousness; cf. ‘idea unit’ in Chafe, 1980). An ‘intonation unit’ contains concepts: the ideas of objects, events, and properties. Such a concept may be in one of three states at any one time: active, semi-active, or inactive. The speaker “makes changes in the activation states of certain concepts during the initial pause, changes which determine the content and form of the following intonation unit” (p. 48). The division into intonation units is not related to the state of the concepts. A previously active concept may then be pronominalized. Active concepts expressing a starting point cannot be pronominalized. Concepts marking a contrastive accent cannot be pronominalized either. Concepts from the semi-active state are referred to as accessible. A concept can become accessible in two ways: when a concept is deactivated, it does not become inactive immediately, but stays in the peripheral memory for some time, it thus remains accessible. The second way is when these concepts belong to the set of expectations associated with a concept in the discourse, the ‘scheme’. Inactive concepts are new. To account for the fact that speakers usually express only one new concept in one idea unit, Chafe introduces the one new concept at a time constraint. A concept can express the starting point of an intonation unit, together with a concept that adds information about this starting point. The light starting point constraint states that a starting point usually is a given concept. The elements described so far are used to mark the structure of intonation units. Above the intonation unit there are more levels: sentences, paragraphs and, ultimately, the narrative. These are described below.

A division into paragraphs (again in the transcription of the spoken discourse) is made through the location of responses from the hearer and through pausal evidence. Sentences are defined by the occurrence of imagined falling pitches, and are independent of the activation states. They are determined by the decision of the speaker to structure the discourse as clearly as possible. The entire narrative can, according to Chafe, be thought of as an island of memory, isolable from the rest of the conversation.

The goal of Chafe’s study was to provide some very general principles that apply to spontaneous spoken language. The universality of these principles,
however, remains to be demonstrated. In assuming a third level of focus (semi-active, active, inactive), this theory goes one step further than the often used binary distinction in given vs. new in the phonetic literature (see, for instance, Nooteboom & Kruyt, 1987). The distinction used by Chafe (1987) is ternary instead of binary, and thus potentially more accurate. Some definitions, however, are not totally clear. For instance, the difference between a starting point and the beginning of a new paragraph is not evident. When do these two coincide and when do they not? Another point is that Chafe does not assume a ‘common ground’, which is present in all listeners’ minds. This common ground is comparable to ‘knowledge of the world’, and can account for the fact that some entities are new in the discourse, but not classified as inactive information, because they are assumed to be generally known.

This theory, in contrast with RST described above, is capable of accounting for the internal structure of clauses. RST is used here to define the distinction between sentence and paragraph boundaries independently of intonation. As indicated above, Chafe’s analysis concerns primarily the activation state of a concept in the mind of a supposed hearer. The analysis, however, still makes some use of acoustic features: clauses are defined as ‘intonation units’, which are detected by ‘pauses’, and sentences are defined by the occurrence of ‘falling pitches’. This means that we will need another theory that is even more accurate in defining the internal structure of clauses, and that is not based on any acoustic feature. Prince’s theory seems to meet these requirements.

2.2.4. Assumed Familiarity

According to Prince (1981), natural language presents an informational asymmetry in that some units seem to refer to ‘older’ information than others. Distinctions in given vs. new information can be found on three levels: in the sentence, in the discourse, and in the discourse model used by the participants. On all levels, the crucial factor seems to be the “tailoring of an utterance by a speaker to meet the needs of the assumed receiver” (p. 224). On the basis of these three levels Prince proposes a model that is applicable to naturally occurring texts in assigning the structure and the distribution of given vs. new information.

In the literature, the given-new distinction is presented under various different names, for instance: given-new, old-new, known-new, presupposition-focus, presupposition-assertion, etc. For an overview, see Lambrecht (1994) and Dryer (1996). However, these notions have never been characterized satisfactorily so that researchers can use them adequately and make them operational. We will present the definitions used by Chafe (1976), Clark &
Haviland (1977), Halliday (1967) and Kuno (1972, 1978), using Prince’s terminology. We will then present the model proposed by Prince to account for the structure and the distribution of given vs. new information. Instead of describing the differences between given information versus new information, as is usually done, Prince distinguishes between three types of givenness: Givenness as predictability/recoverability (givenness\(_p\)), Givenness as salience (givenness\(_s\)), and Givenness as shared knowledge (givenness\(_k\)). These types are discussed below, and the different definitions used in the literature are integrated in this tripartition.

- **Givenness\(_p\) as predictability/recoverability.** “The speaker assumes that the hearer can predict or could have predicted that a particular linguistic item will or would occur in a particular position within a sentence” (p. 226).

Kuno (1978) defines old-new in terms of recoverability: “an element in a sentence represents old, predictable information if it is recoverable from the preceding context; if it is not recoverable, it represents new, unpredictable information” (pp. 282-283). Halliday (1967) defines given-new differently: given is defined as “the complement of [an intonationally] marked focus” (p. 208). New information is “information ... that the speaker presents ... as not being recoverable from the preceding context” (p. 204). Halliday & Hasan (1976) define given as “expressing what the speaker is presenting as information that is recoverable from some source or other in the environment - the situation or the preceding context” (p. 326). It is not clear what this ‘source’ exactly is. Kuno’s predictability looks similar to Halliday’s recoverability, but what is old for Kuno is not necessarily given for Halliday. Prince proposes a principle that could be included in the predictability of Kuno, the Parallelism Principle: “a speaker assumes that the hearer will predict, unless there is evidence to the contrary, that (a proper part of) a new (conjoined?) construction will be parallel/equivalent in some semantic/pragmatic way(s) to the one just processed” (p. 228). Prince concludes that it is crucial to consider the speaker’s hypotheses about the hearer’s beliefs and assumptions in the notion of givenness.

- **Givenness\(_s\) as salience.** “The speaker assumes that the hearer has or could appropriately have some particular thing/entity ... in his/her consciousness at the time of hearing the utterance” (p. 228).

This definition represents the theory of Chafe (1976). Chafe (1976) defines given as “that knowledge which the speaker assumes to be in the consciousness of the addressee at the time of the utterance” (p. 30) and new as “what the speaker
assumes he is introducing into the addressee’s consciousness by what he says” (p. 30). This presents a binary distinction. Furthermore, a given element must have an explicit referent in the discourse.

- **Givenness as shared knowledge.** “The speaker assumes that the hearer ‘knows’, assumes, or can infer a particular thing (but is not necessarily thinking about it)” (p. 230).

Clark & Haviland (1977) define given as “information [the speaker] believes the listener already knows and accepts as true” and new as “information [the speaker] believes the listener does not yet know” (p. 4).

Kuno (1972) introduced the notions of anaphoric and non anaphoric. These also fall under the term of givenness. An element is anaphoric, if “[its] referent has been mentioned in the previous discourse” or is “in the permanent registry” (p. 270), which refers to what the speaker assumes about the hearer’s assumptions. This is related to the tendency to put old information before new information, old referring to shared knowledge.

How do these three types of givenness relate to each other? The three types are not mutually independent. Ultimately, all levels refer to extra-linguistic phenomena. The understanding of the givenness as predictability or salience is dependent on the understanding of the givenness in the sense of shared knowledge.

In the actual model proposed by Prince (1981), ‘shared knowledge’ is replaced by ‘assumed familiarity’. The knowledge and assumptions of the speaker and the hearer are important insofar as they affect the forms and understanding of linguistic productions. Three parts are needed in the model: linguistic form, values of assumed familiarity and the correlation between these two. Prince describes the model by comparing a text to a recipe: the text presents a “set of instructions from the speaker to the hearer on how to construct a particular discourse model” (p. 235).

A new entity in the discourse can be **brand new** (cf. to be bought in a store) or **unused** (cf. to be taken from a shelf). The brand new entities can be **anchored** (linked by means of another NP to some other entity) or **unanchored**. All anchored entities contain at least one anchor that is not a brand new item itself. The distinction between brand new and unused can be related to the linguistic representation of these items, i.e. indefinite versus definite NPs. This means that indefinite NPs are classified as brand new, while definite NPs are usually
NPs which are already present in the discourse are presented as *evoked* entities. Items can be *textually* evoked, meaning that at one point in the discourse this item was new, or *situationally* evoked, meaning that the hearer assumes that the listener can evoke it by himself, from the situation.

The third type are the *inferrable* entities. An entity is inferrable if the speaker assumes that the hearer can infer it from entities already evoked in the discourse or from knowledge of the world. These are called *noncontaining*. *Containing* inferrables form a special subclass of inferrables: "what is inferred off of is properly contained within the inferrable NP itself; [...] one of these eggs is a containing inferrable, it is inferrable, by set-member inference, from these eggs which is contained within the NP and which, in the usual case, is situationally evoked" (p. 236).

Figure 2.3 presents the different discourse entities. The terms placed between parentheses are additional to the default notion: ‘evoked’ generally means ‘textually evoked’, just as ‘inferrable’ generally means ‘noncontaining inferrable’, ‘brand new’ generally means ‘brand new unanchored’.

![Taxonomy of Assumed Familiarity by Prince (1981)](image)

The textually evoked items can be further divided into *remote* (or ‘displaced’) and *current* (Brown, 1983). The remote textually evoked items are too far back in the discourse to be pronominalized (cf. semi-active in Chafe’s theory), for instance by the presence of some kind of discourse boundary, while the current
textually evoked items can reoccur in the form of a pronoun (cf. active in Chafe’s theory).

Prince’s taxonomy is used by a number of researchers to relate the information structure of discourse to the linguistic (Birner, 1994a; 1994b; to appear) or phonetic realization by the speaker (Brown, 1983; Caelen-Haumont & Bessac, 1997).

Related to the issue of given and new information is the distinction between ‘topicality’ and ‘focality’. Topicality can be seen as “the things we talk about” (Dik, 1997, p. 310), and focality as “the most important or salient parts of what we say about the topical things” (op. cit., p. 310). Topics may be introduced to and maintained in the discourse in different ways, and according to different strategies (for instance a ‘sub-topic’, which is comparable to the category ‘inferrable’ in Prince (1981)): if one is talking about a party, it is reasonable to talk about ‘music’ (example from Dik, 1997). These ‘topicality strategies’ concern such questions as how new topics are introduced in the discourse, how topics are kept in the discourse as given topics, what sub-topics can be associated with given topics, and what strategies are available to reintroduce given topics as so-called resumed topics. Focal information is not necessarily always equal to new information. Information that is already available may be focused on in order to emphasize it. Focality may be expressed in one or more ways: prosodic prominence, special constituent order, special focus markers, special focus constructions. Focus is, as explained above, not equal to accent, nor to prominence. Prosodic prominence (often expressed by a pitch accent) is one means available to the speaker to draw the listener’s attention to certain information. A detailed account of ‘topicality’ and ‘focality’ can be found in Dik (1997, ch. 13).

2.2.5. Summarizing remarks and proposed analysis

The division into three basic parts used by Prince (1981) is roughly comparable to the division used by Chafe (1987). The tripartition used by Chafe is less specific. Prince’s new items coincide fully with Chafe’s inactive concepts, but are subdivided, and thus more subtle. The semi-active concepts of Chafe coincide with Prince’s inferrable, but include also the remote textually evoked items. In Chafe’s theory only given or active concepts can be pronominalized. This indicates that the remote textually evoked items are not available for pronominalization, probably because a paragraph boundary occurs between the original item and the evoked item. This boundary blocks the pronominalization.
We assume that such a boundary coincides with the paragraph boundaries found in the RST analysis.

One fundamental difference between the two analyses, as already indicated, is that Chafe’s analysis concerns the activation state of a referent in the mind of the hearer, whereas Prince’s analysis refers to the formulation chosen by the speaker: a referent is classified as ‘brand new’ if the speaker formulates this referent as ‘brand new’. The analysis proposed by Prince seems more accurate in distinguishing several levels, thus assuming a hierarchical structure. Furthermore, Prince’s analysis is based on the linguistic representation of elements, and thus seems more suitable to indicate focal structure without making use of acoustic features than the analysis proposed by Chafe. The Prince analysis, however, does not apply to verbs. Verbs are said to be part of the concept expressed by the noun (e.g. ‘reading a book’ counts as one concept), and are therefore not classified separately.

In Chafe’s analysis, adverbial constituents can be classified as certain ‘orientations’.

The following preliminary approach to analyze discourse structure is formulated on the basis of the findings presented above. The applicability of this method, which we called ‘Information Structure In Discourse’, was tested in a pilot experiment (section 2.4), and subsequently adjusted.

I. Discourse structure. The division of the discourse into smaller units is mainly inspired by RST. We distinguish the following functional pieces: clauses, sentences, and paragraphs. A clause is defined as a unit containing words or word groups (concepts) taken together on semantic, syntactic, or functional grounds. This means that a clause can consist of a verb plus nouns and determiners or modifiers, but also of an idiomatic expression. Contrary to RST, we consider main clauses and subordinate clauses as separate units, since they can be formed on a semantic, syntactic, or functional ground. Our clauses can be compared to what Chafe (1987, 1994) calls ‘intonation units’, but our clauses are evidently not based on any intonational features, so as to avoid the circularity mentioned in the introduction. In RST, clauses are the only units of analysis; no other unit is defined between the clause and the whole discourse. This distinction evidently does not cover all possible subunits within a (spoken) discourse.2 In written text, one can also distinguish sentences and paragraphs. Therefore, we also distinguish sentences and paragraphs. (This distinction is thus additional to the clauses ‘taken from’ RST.) A sentence may contain

---

2 We will not discuss the issue of how to define possible subunits within the discourse.
one or more clauses, which together form a functional unity. Thus, a main clause with its subordinate clauses typically form a sentence. A paragraph consists of one or more sentences dealing with the same topic. These definitions are not those proposed by Chafe, since they are based on acoustical cues rather than on independent grounds.

II. Information structure. As for a subdivision into concepts, we mainly follow Prince’s analysis. Nominal constituents can be classified as follows, using so-called ‘textual labels’: A brand new (bn) element refers to information that is completely new to the addressee. This addressee can be either the reader of the transcription or the listener to the spoken version of the discourse. Both types are used in our research. Brand new elements are usually indefinite nouns or generic expressions. The label brand new anchored will not be used in our method; since we did not find clear examples of this label in our text, it did not seems necessary to maintain this label. 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. The label containing inferrable is not included in our method; again we did not find clear examples of this label. Evoked elements have already been mentioned in the discourse. They can be 1) textually evoked (et): the noun is evoked by a real pronoun, 2) 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 being used to avoid ambiguity, 3) situationally evoked (es): the referent of a noun or pronoun can only be found in an extra-textual context. Prince’s theory does not include any labels for elements like verbs or for adverbs. Since these elements can (and probably do) express valuable information as well, we propose some additional labels. Adverbs or other adverbial expressions of time or place (not sentence initial) will be labeled modifier (mod), and can contain new information depending on the context. The term ‘modifier’ is used here as a pragmatic label rather than a semantic one. Sentence initial adverbs or adverbial expressions are labeled orientation (or). The label modifier is thus introduced as a new label. If verbs are used as nominalized verbs, they can easily be classified as nouns. Verbs are thus classified using the labels unused, inferrable and evoked in the same way as for nominal constituents. Since verbs can intuitively not be brand new, they are thus labeled as unused. Distinguishing between textually or situationally evoked verbs is not always easy, and for our
research purposes it is not of crucial importance. Verbs will thus only be
classified as evoked. The verb phrase as a whole is labeled, the auxiliary
and the main verb thus being considered as a unitary concept. Prepositions
which are part of a verb are included in the concept as well. This implies
that the total number of concepts per discourse is smaller than the total
number of words.

2.3. Pilot experiment: Perceived prominence

2.3.1. Spoken material and textual analyses

A listening experiment was carried out to investigate whether there is perceptual
evidence for the proposed ‘Information Structure In Discourse’ analysis, in other
words, the question ‘What is the relation between the perceptual judgements of
listeners and the textual analysis?’ was asked. What ‘extra’ cues does prosody
add to the textual structure of a discourse? In this pilot experiment, we will focus
on the local level, i.e. the information structure, rather than on the global
discourse structure. This is done since the division of the individual concepts is
much more delicate than the more straightforward division of a discourse into
smaller subunits, and should therefore be well chosen.

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 story telling situation. These retold versions were transcribed literally by
the author of the present study, including all hesitations and false starts, but
without any punctuation marks or capitals. The next day the verbatim
transcription was read aloud by the same speakers (‘re-read version’), every
speaker reading his/her own story. Each speaker was encouraged to read the text
carefully before reading it aloud, to mark punctuation, and to correct hesitations
or false starts. This was explicitly not done by the transcriber, since any change
or mark in the text might 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 with a Sennheiser MKH105
high frequency condenser microphone, and stored on DAT-tape. The speakers
participated on a voluntary basis.
The transcribed retold versions were subject to a discourse analysis done by five experts by means of the ISID method described in section 2.2.5. First, the discourse was divided into clauses, sentences, and paragraphs. Next, all nominal and verbal elements were given a label according to the information status they expressed, applying the labels \(bn, u, i, et, etd, es,\) and \(mod.\) Also, ‘orientations’ were labeled, using the label \(or.\) This was done first of all by the author of the present study. The main difficulties encountered in analyzing the texts were subsequently discussed in a plenary session with a panel of the five text analysts, all familiar with discourse theory. Where necessary, the originally proposed analyses were adapted, following the suggestions of the analysts. Generally, one common opinion could be formulated. In Figure 2.4 an example of a text analysis is given. Clauses are numbered 1 through 11. Since we were mainly interested in the local information structure rather than in the global discourse structure, sentences and paragraphs are not indicated. English glosses are given after each clause in quotes. Nominal and verbal constituents are placed in parentheses, the matching label is attached by means of a ‘-‘. Parts of constituents separated within the clause are connected with subscript ‘1’. The labels are coded as follows: \(or = \) orientation, \(bn = \) brand new, \(u = \) unused, \(i = \) inferrable, \(et = \) evoked textually, \(etd = \) evoked textually displaced, \(es = \) evoked situationally, \(mod = \) modifier.

### 2.3.2. Judging prominence

We hypothesized that all information that is was some way ‘new’ to the discourse, would be perceptually judged as prominent (labels \(bn, u, i, mod,\)) while the information that was ‘already mentioned’ would not be perceptually judged as prominent (labels \(or, et, etd, es).\) With this in mind we compared the two speaking styles ‘retold’ and ‘re-read’, as well as possible differences between male and female speakers.

Since we are interested in the perceived structure of spoken discourse, and since we want to be able to compare the results for different speaking styles, we only used the retold and re-read versions in this experiment, since these are in principle lexically identical; the original read aloud version was therefore left out of consideration. This resulted in 16 different texts (8 speakers x 2 versions). These were randomly ordered in such a way that every listener evaluated three different texts and each text was graded by three different listeners. The 16 listeners were seated in individual booths in the language laboratory of the Faculty of Humanities of the University of Amsterdam, and listened to the speech material over headphones. The verbatim transcription of the spoken text
was used as an answer sheet. All 16 listeners were students or staff members of the University of Amsterdam. Student listeners were paid for their participation.

   ‘it is about two people’
2. die [wonen-u] [in de stad-u]  
   ‘who live in the city’
3. en [op een morgen-or] [worden] [ze-et] [wakker-u]  
   ‘and on one morning wake they up’
4. en [dan-or] [zien-u] [ze-et] dat het [heel hard-mod] [gesneeuwd-u] [heeft-i]  
   ‘and then see they that it very heavily snowed has’
5. [het-es] is dus [een verhaal-bn] [in de winter-i]  
   ‘it is thus a story in the winter’
   ‘and they decide to that day once in the woods to go see’
7. hoe [het-et] er dan [daar-et] [uit ziet-i]  
   ‘how it there then looks’
8. [de stad-etd] uit [het bos-etd] in  
   ‘the city out the woods in’
9. [in het bos-etd] is het eeh [heel heel dik-mod] [besneeuwd-e]  
   ‘in the woods is it very very heavy snow-covered’
10. [de takken van de jonge bomen-i] die [buigen-u] over  
    ‘the branches of the young trees they bend over’
11. en [daar-et] moeten [ze-et] [soms-mod] [onderdoor kruipen-u] ...  
    ‘and there must they sometimes underneath crawl’ ...

Figure 2.4  Example of a text analysis of part of a retold story in Dutch. English glosses are given in quotes.

The listeners were instructed to evaluate the spoken versions in terms of prominence, using the speech signal as anchor point. Each listener was presented with an individual tape containing four different spoken versions of the story, each version either a retold version or a re-read transcription. A fourth speaker (identical for all listeners, and different from all other speakers) was added to the three ‘real’ speakers as an exercise at the beginning of the tape. This additional speaker was not included in the results. The listeners 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, thus 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). The two-hour time period given to fulfill the task was sufficient for all listeners.

2.3.3. Results

Apart from the one jointly-analyzed text, each text was evaluated by three different listeners. For each of the 16 verbatim transcriptions the analysis based on the text alone was taken as reference point; the text of the re-read versions only differed on some minor points from the verbatim transcription of the retold story. The perceptual judgements were compared to these analyses. For every text, version, and listener a correspondence matrix was made, in which the labels from the text analysis were matched against the prominence judgements 1, 2, and 3. This resulted in 48 matrices (8 speakers x 2 versions x 3 listeners per text).

'Zero labels' were added to provide for the cases in which a word was underlined but no specific judgement was given (zero perception, 'X'), and those in which a word was underlined that did not have a proper label in the text analysis (zero text analysis, '0'). As formulated above, we have very clear predictions about the relation between textual label and perceived prominence. Therefore, words with no textual label or prominence judgement can be left out of consideration, since they fall outside the scope of the present experiment.

For a first impression of how the perceptual judgements of prominence might be related to the textual analyses, we normalized the judgements to percentages (spontaneous and re-read versions taken together) by setting all given judgements to 100%, and added up all judgements (Table 2.1). The three perceptually most relevant labels are unused (22% of all judgements), brand new (17% of all judgements) and modifier (16% of all judgements). 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.

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 (‘evoked textually displaced’), 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’ (as a resumed topic), 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 using the full noun ‘the forest’ to avoid ambiguity.

Table 2.1 Perceptual prominence judgement matched against textual label, normalized to percentages, for all 8 speakers and all 16 listeners and both versions together. Total percentages are given for each label separately as well as for the category as a whole. The mean number of textual labels per category (standard deviations between parentheses), averaged over the eight speakers, are given in the last column.

Prominence levels X, 1, 2, and 3 refer to the relative prominence judged by the listeners; the textual labels 0, or, mod, bn, u, i, etd, et, and es refer to the different categories of information from the discourse analysis. Lines separate the different categories.

<table>
<thead>
<tr>
<th></th>
<th>X = no level</th>
<th>1 = much emphasis</th>
<th>2 = less emphasis</th>
<th>3 = little emphasis</th>
<th>Total %</th>
<th>Mean no. and sd</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>0.04</td>
<td>1.50</td>
<td>2.16</td>
<td>1.81</td>
<td>5.52</td>
<td>7.34 (9 (4))</td>
</tr>
<tr>
<td>or</td>
<td>0.00</td>
<td>0.51</td>
<td>0.96</td>
<td>0.34</td>
<td>1.82</td>
<td></td>
</tr>
<tr>
<td>mod</td>
<td>0.06</td>
<td>4.34</td>
<td>6.96</td>
<td>4.25</td>
<td>15.61</td>
<td></td>
</tr>
<tr>
<td>bn</td>
<td>0.00</td>
<td>5.14</td>
<td>7.69</td>
<td>4.55</td>
<td>17.37</td>
<td>55.47 (98 (16))</td>
</tr>
<tr>
<td>u</td>
<td>0.08</td>
<td>6.43</td>
<td>9.95</td>
<td>6.03</td>
<td>22.49</td>
<td></td>
</tr>
<tr>
<td>i</td>
<td>0.04</td>
<td>3.97</td>
<td>6.27</td>
<td>3.73</td>
<td>14.01</td>
<td>14.01 (44 (6))</td>
</tr>
<tr>
<td>etd</td>
<td>0.05</td>
<td>4.07</td>
<td>6.39</td>
<td>3.81</td>
<td>14.32</td>
<td></td>
</tr>
<tr>
<td>et</td>
<td>0.04</td>
<td>1.59</td>
<td>3.95</td>
<td>2.59</td>
<td>8.16</td>
<td>23.18 (74 (15))</td>
</tr>
<tr>
<td>es</td>
<td>0.00</td>
<td>0.24</td>
<td>0.20</td>
<td>0.26</td>
<td>0.70</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>0.31</td>
<td>27.78</td>
<td>44.54</td>
<td>27.36</td>
<td>100%</td>
<td>100% (225 (38))</td>
</tr>
</tbody>
</table>

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%) consists of the items orientation (or) and zero text analysis judgements (0). Zero judgements refer to concepts judged as perceptually prominent (the word was underlined), but no textual label was given.

When looking at the relative prominence judgements (1, 2, or 3, see last row in Table 2.1), 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 (X). 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, which may indicate that only in extreme cases was a 1 or a 3 judged. Therefore, in the following, we will take into account only the total percentage of prominence judgements per textual label and category, as given in the ‘Total %’ column of Table 2.1.

Research by Blaauw (1995) made clear that in Dutch, the speaking styles retold and re-read are quite distinct, both acoustically and perceptually: listeners are able to classify small stretches of speech (2 to 6 syllables) above chance as either spontaneous or read. Our concern is, of course, whether this distinction also affected the listeners in their prominence perception. In other words, did they perceive prominence differently in the two speaking styles? If so, what categories of information are concerned?

Table 2.2 Overall percentage prominence judgements, broken down for speaking style and sex of speaker. Labels belonging to the same category are separated by a line; the category name is given in the first column (‘Category’).

<table>
<thead>
<tr>
<th>Category</th>
<th>Textual label</th>
<th>Speaking style</th>
<th>Sex of speaker</th>
</tr>
</thead>
<tbody>
<tr>
<td>Disc. mrkr</td>
<td>0</td>
<td>Retold: 4.6</td>
<td>Male: 5.4</td>
</tr>
<tr>
<td></td>
<td>or</td>
<td>Re-read: 6.6</td>
<td>Female: 5.9</td>
</tr>
<tr>
<td>New</td>
<td>mod</td>
<td>Retold: 15.6</td>
<td>Male: 13.2</td>
</tr>
<tr>
<td></td>
<td>bn</td>
<td>Re-read: 17.1</td>
<td>Female: 19.3</td>
</tr>
<tr>
<td></td>
<td>u</td>
<td>Retold: 23.0</td>
<td>Male: 23.4</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Re-read: 22.0</td>
<td>Female: 21.7</td>
</tr>
<tr>
<td>Inferrable</td>
<td>i</td>
<td>Retold: 14.4</td>
<td>Male: 14.7</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Re-read: 13.4</td>
<td>Female: 13.1</td>
</tr>
<tr>
<td>Evoked</td>
<td>etd</td>
<td>Retold: 15.1</td>
<td>Male: 14.5</td>
</tr>
<tr>
<td></td>
<td>et</td>
<td>Re-read: 13.2</td>
<td>Female: 13.7</td>
</tr>
<tr>
<td></td>
<td>es</td>
<td>Retold: 8.1</td>
<td>Male: 8.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Re-read: 9.1</td>
<td>Female: 9.0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Retold: 0.7</td>
<td>Male: 0.6</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Re-read: 0.7</td>
<td>Female: 0.8</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td>100%</td>
<td>100%</td>
</tr>
</tbody>
</table>
The first two data columns of Table 2.2 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 being significant ($\chi^2=18.52$, df=8, p<0.05). We expected larger differences between the two speaking styles, since they are perceptually quite distinct (Blaauw, 1995). This was confirmed in a small classification experiment, in which 8 male and 8 female students and a number of staff members of the Institute participated. The listeners were asked to classify medial fragments of 1.5 minutes 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 in 60% of the cases were 'very sure' of their choice. This suggests that also the listeners in the evaluation experiment, who listened to the entire spoken text, should have been able to hear the difference in speaking style. Therefore, the fact that the two speaking styles are perceptually distinct seems not to affect the listener's perception of prominence, at least in our experiment. The large differences found by Blaauw (1995) between the two speaking styles concern the ability of listeners to classify stretches of speech as spontaneous or read speech. The ability of listeners to evaluate one specific style on, for instance, the occurrence of prominence, was not investigated in that study.

We do find, however, that whenever the retold speaking style dominates in number of prominence judgements, this is for the major categories from Table 2.1 ('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. Since the two speaking styles do not differ significantly, they can be taken together in the rest of the experiment.

The material presented to the listeners was produced by four male and four female speakers. Research by Tielen (1992) has shown that listeners are almost unambiguously able to distinguish male from female voices in different speaking styles. We want to know whether the sex of the speaker is related to the perception of prominence. In other words, is there a difference in the way male and female speakers are treated by the listeners? In the following, we will thus look for possible differences between prominence perception in the two speaking styles retold and re-read, and between the ways in which the discourse produced by male and female speakers are perceived.

The last two columns of Table 2.2 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 ('brand new', 'unused', 'inferrable', and 'evoked textually displaced'), the male
speakers were given higher prominence scores than the female speakers. The female speakers, however, emphasized more modifiers than did the male speakers. The effect of sex of speaker is significant ($\chi^2=36.3$, df=8, $p<0.01$).

Finally, something should be said about the so-called ‘zero judgements’. Overall, they provide for 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. When examined more closely, these items 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 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 unit. However, the auxiliary (as any word in the text) can be emphasized for contrastive reasons. The informational status of a concept 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. Within the ISID framework, this type of information is labeled as ‘evoked’, since the change of information can in fact only be made when listening to the text.

2.3.4. Summarizing remarks

In this section some preliminary conclusions concerning the data presented above are discussed. The method analyzing 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 not yet strict enough to be used as working definitions of types of information status. Within the collection of nine textual labels, clusters can be defined of labels behaving identically in the analysis, which indicates that matters can be simplified further. 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 Tables 2.1 and 2.2. 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. These will thus constitute the category discourse markers.
Both Tables 2.1 and 2.2 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 no longer 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. Prince groups together *inferrables* with 'Discourse-old Nonpronominals' (our *evoked textually displaced* items).

The final version of our framework 'Information Structure In Discourse' is presented in the next section.

### 2.4. Current framework: *Information Structure In Discourse*

First, some preliminary remarks. It is important to note that the current framework is *not* a mere collection of elements and/or items from the theories mentioned above. None of the above-mentioned theories offered a clear and unambiguous approach to analyze discourse structure at the levels needed for our investigation. The current framework thus has a status of its own, and does not aim to improve or elaborate any of the theories. Neither does it claim to be a full theory of discourse analysis. On the basis of the different theories and approaches found in the literature, we developed a method suitable to analyze our (spontaneous) data. This method can, of course, be applied to other material, whenever one wishes to investigate the relation between discourse structure and prosodic aspects.

Recapitulating, the application of the framework is as follows: It consists of two parts: a 'global' level and a 'local' level of structure. In Table 2.3 these two levels are separated by a dashed line. The starting point of the analysis is a verbatim transcribed spoken text, without any punctuation and/or typographical lay-out. The first step is to identify the *hierarchical global level of discourse structure* (going from the largest to the smallest functional unit): discourse → paragraphs → sentences → clauses → word groups (i.e. concepts). The
The next step is to determine the local level of information structure. For each concept in the clause, the information status is labeled accordingly. In the previous section four main category labels were presented: new, inferrable, evoked, and discourse marker. In the final analysis presented here, some additional categories are included in order to be able to test specific hypotheses concerning accent placement (see chapter 5). The modifiers included in the category new are labeled as a separate category. Furthermore, a separate category connectors is introduced, as well as a category verbs. The category connectors is necessary to account for the fact that in spontaneous speech almost all clauses begin with 'and ...'. Since we want to investigate the prosodic marking of discourse structure in spontaneous speech, these characteristic elements should be available. As mentioned above, verbs are labeled in the same way as nominal elements. This means that we cannot see whether a prominently perceived new item is a nominal or a verbal element. As was explained, Prince does not treat verbs separately, but classifies them as part of the concepts. At this point, however, we feel that it is necessary to create a verb-category, precisely to see whether verbs are indeed not a separate category. Therefore, the labels new, inferrable, and evoked will be exclusive for nominal elements, and all verbal elements will be labeled as verbs, with no specific information label. Thus, in total we have seven labels to mark local information structure in discourse: new, inferrable, evoked, modifier, discourse marker, connector, and verb. Table 2.3 presents an overview of the framework: for each label the function in the
analysis is given. The functions described in Table 2.3 for the labels *new*, *inferrable*, and *evoked* are adapted from Prince (1992).³

The taxonomy of Assumed Familiarity is ‘rephrased’ by Prince in terms of the point of view of the *discourse model*, instead of from the point of view of the *hearer’s mind*. Prince (1992) makes a four-way distinction in the following way:

<table>
<thead>
<tr>
<th>Hearer-new</th>
<th>Discourse-new</th>
<th>Discourse-old</th>
</tr>
</thead>
<tbody>
<tr>
<td>brand new</td>
<td>unused (inferrable?)</td>
<td>evoked</td>
</tr>
</tbody>
</table>

Note that the category *inferrable* is not included here. The slot ‘discourse-old & hearer-new’ is intuitively empty. However, as Hirschberg (1997) has pointed out, this situation can occur when the listener has simply not understood the speaker, and asks ‘Could you say that again, please?’

The results of our pilot experiment suggest that *brand new* and *unused* can be classified as the same category, at least for the purpose of our research. Furthermore, following the definition of *inferrable* as presented earlier, it can very well replace the *unused*-slot in the diagram above. If we do this, the distinction by Prince (1992) perfectly matches the distinction that followed from the results of our pilot experiment, where three major categories for nominal elements are distinguished (new, inferrable, and evoked).

³ The taxonomy of Assumed Familiarity is ‘rephrased’ by Prince in terms of the point of view of the *discourse model*, instead of from the point of view of the *hearer’s mind*. Prince (1992) makes a four-way distinction in the following way:
Table 2.3. Model of ‘Information Structure In Discourse’. For each label, the level at which it is applied (global or local, separated by a dashed line) as well as its function in the analysis is given. An example of a text analysis is given in Figure 2.3, where the model is applied to the verbatim transcription of a retold text.

<table>
<thead>
<tr>
<th>Level</th>
<th>Label</th>
<th>Function in analysis</th>
</tr>
</thead>
<tbody>
<tr>
<td>GLOBAL</td>
<td>I, II, III, ...</td>
<td>division into paragraphs</td>
</tr>
<tr>
<td></td>
<td>a, b, c, ...</td>
<td>division into sentences</td>
</tr>
<tr>
<td></td>
<td>1, 2, 3, ...</td>
<td>division into clauses</td>
</tr>
<tr>
<td></td>
<td>various codings</td>
<td>individual concepts</td>
</tr>
<tr>
<td></td>
<td>new</td>
<td>discourse new &amp; hearer new</td>
</tr>
<tr>
<td></td>
<td>inferrable</td>
<td>discourse new &amp; hearer old</td>
</tr>
<tr>
<td>LOCAL</td>
<td>evoked</td>
<td>discourse old &amp; hearer old</td>
</tr>
<tr>
<td></td>
<td>modifier</td>
<td>clause internal adverb of time, place, ...</td>
</tr>
<tr>
<td></td>
<td>discourse marker</td>
<td>clause initial adverb of time, place, ...</td>
</tr>
<tr>
<td></td>
<td>connector</td>
<td>connecting element ‘and’ between clauses</td>
</tr>
<tr>
<td></td>
<td>verb</td>
<td>verbal elements</td>
</tr>
</tbody>
</table>

Figure 2.5 gives an example of a discourse analysis. The framework is applied to the verbatim transcription of a retold text, as spoken by one of the speakers. It is the same fragment as presented in section 2.3.1. The original Dutch text is given, with an English translation. The original Dutch text on which the retelling is based and its English translation are given in Appendix A and Appendix B, respectively.
Het gaat over twee mensen die wonen in de stad. En one morning they wake up and then they see that it has snowed very heavily. So it is a story in wintertime. And they decide to go and see in the woods, to check how it looks over there. Out of the city, into the woods. In the woods there is a whole lot of snow. And the branches of the trees are hanging. And sometimes they have to crawl underneath them.

2.5. Conclusions

Inspired by a number of theories on text structure found in the literature, we have developed a model that should be able to analyze both written and spoken discourse on several levels. First of all, we define a global structure in terms of the division into paragraphs, sentences, clauses, and concepts, referred to as the discourse structure. Secondly, we define a local structure in terms of determining the information structure of the individual concepts (new, inferrable, evoked, discourse marker and connector, modifier, and verb), referred to as the information structure. The method was designed to create a reference point to which prosodic aspects of the spoken discourse (boundary marking and accent...
placement) could be related. Therefore, the model needed to be strictly textbased, or prosody-independent, to avoid circularity.

The results of a pilot experiment, and the subsequently made adjustments, led us to propose the model of ‘Information Structure In Discourse’ (ISID). This model was used to analyze (the transcriptions of) the spoken texts by the eight speakers in our study, on the levels of discourse structure mentioned above, thereby providing the larger part of all words occurring in the text with a limited set of labels. The judgements made by naive listeners can be described in terms of these labels, which indicates that the model is very suitable to analyze the relation between textual structure and perceived prominence in discourse.

The next chapter will give a detailed description of the speakers and the discourse they produced in this study. Speaker-specific discourse features will be presented and discussed, as well as differences and similarities on an overall acoustic level in terms of temporal aspects, such as speech tempo and pausing strategy, and intonational aspects, such as mean fundamental frequency, range, and the realization of peak and end tones.