How to specify focus without using acoustic features

Monique E. van Donzel

Abstract

In this paper we present first of all an overview of the literature on the subject of focus
and discourse structures. Within the theories about focus we make a distinction between
the definitions of focus based on acoustic features and those not based on acoustic
features. We further present a method that should be able to indicate the internal focal
structure of a text, both in a read aloud version and in a freely retold version. This
method is illustrated by an example text analysis.

1. Introduction

In our four-year project *Acoustic-phonetic correlates of focusing in discourse and
dialogue* (Koopmans-van Beinum 1994) we will investigate the possible ways in
which a speaker can put words or word groups in focus. *Focus* is usually defined by
means of *intonation*, namely by stating that a word or a word group is placed in focus
if an *accent* is realized on that word or word group. This kind of definition, however,
may lead to circularity: the possible acoustic features are already included in the
definition itself. This circularity can be avoided by looking for a way to define focus
without including any intonational features. Therefore, it is necessary to define the
notion of focus, as well as other notions related to this, such as *old* vs. *new*
information. In spoken texts focus is realized by the presence or absence of accents. In
written texts, words can also be perceived by the reader as being more or less
important (plus or minus focus). In that case, however, there is evidently no relation
with accents. The degree of importance can then best be indicated using for instance
the terms new and old information.

1.1. Organisation of the paper

In this paper we will give an overview of the literature concerning the focal structure
of texts. We will distinguish between approaches in which the focal structure of a text
is related to acoustics, and approaches in which it is not acoustically defined.

First of all, we will describe some definitions of focus based on intonation, and
show that these methods cannot be used to define focus in an operational way. These
approaches usually make a very rough distinction between given and new information
only. We will describe the definitions used by Eady & Cooper (1986), Nooteboom &
Kruyt (1986), Horne (1991a, 1991b), and Fowler & Housum (1987).

Secondly, we will present some approaches in which the focal structure of a text is –not based on acoustic features. These are theories about how to present the structure of discourses, based on textual analysis rather than on acoustic measurements. We will describe Rhetorical Structure Theory developed by Mann & Thompson (1988), which results in a very global structure of a text. The approach used by Chafe (1987) is more specific and can account for the focal structure of sentences within a text, as can the approach used by Prince (1981). This last approach is even more detailed than the one used by Chafe.

Finally we will present a method which will be able to detect the focal structure of a text, without making use of acoustic features.

In the next section we will present an abstract of our four-year project (Koopmans-van Beinum 1994).

1.2. Abstract: The acoustic-phonetic correlates of focusing in discourse and dialogue

The structure of information in written texts usually becomes clear by the use of typographical means. In spoken texts it is generally assumed that the speaker may use various acoustic means to assign structure. It is, however, not clear whether this is done systematically. This project concerns two questions. Firstly we concentrate on the acoustic parameters of focusing (intonational, durational and spectral aspects). We will investigate the way in which the speaker marks which words or word groups are important. We will look at possible differences between the way this is done in a (monologue) discourse and in a dialogue, and at possible differences between spontaneous speech and texts read aloud.

Secondly we focus on the perception. We want to find out which correlates are most important for the listener to determine the structure of spoken texts.

One of the questions is whether there is any systematicity in the way speakers use acoustic cues to mark focus in their discourse (cf. Koopmans-van Beinum 1992a; 1992b). The transitional segments (Redeker 1992) are indicated by linguistic markers. It is not known what the acoustic correlates of these markers are, nor how they are marked compared to focus words. These issues will be investigated in our own research project.

2. Theories in which the focal structure is related to acoustic features

2.1. General remarks

The distinction mostly used by phoneticians in structuring information in a text or a discourse is the distinction between old and new information. The texts or discourses used in that sort of research are generally not coherent texts like for instance stories, but rather intended combinations of utterances. Fowler & Housum (1987), however, used monologues. *Old* or *given* information roughly refers to information already presented to the listener in an earlier stage of the discourse, while *new* information refers to material not yet presented to the listener, thus previously unknown. This division is used in various phonetic experiments, designed to determine the acoustic features of old versus new information. The ways in which old and new are defined

differs in various experiments. We will begin by describing some definitions of old/given vs. new information as defined in these experiments, and the problems arising from these approaches.

2.1.1. Eady & Cooper (1986)

Eady & Cooper (1986) define focus following Chomsky (1971), Ladd (1980) and Selkirk (1984). The claim is that different focus scopes are acoustically and perceptually distinct, and are mainly manifested by different intonation patterns and differences in duration. This implies that focus is defined through prosody: a word is said to be in focus if that word bears an acoustically realized accent.

2.1.2. Nooteboom & Kruyt (1986)

Nooteboom & Kruyt (1986) proceed in much the same way. They claim that speakers place words or word groups in focus by means of an accent on that word or on the prosodic head of that word group. The difference between the approaches of Eady & Cooper and Nooteboom & Kruyt is that in the former focus is defined as a property of a word, while in the latter larger constituents can be focused as well. Accents thus mark a constituent as [+ focus]. If a constituent does not have an accent, it is marked [- focus]. This is related to the status of the information expressed: plus focus generally refers to newness of the information, while minus focus refers to givenness. Each plus focus domain is marked by a single accent, while minus focus domains contain no accent. New versus given is defined contextually. Given information generally indicates that the information has already been mentioned by the speaker earlier in the discourse. All the other information is new.

2.1.3. Horne (1991a)

Horne (1991a) defines new as 'brand new' and given as 'mentioned previously'. Again, the status of the information is defined on the basis of context. This relates to focus in that new information is accented, while given information is not. Horne made use of the results from Eady & Cooper (1986).

2.1.4. Fowler & Housum (1987)

Fowler & Housum (1987) make a distinction between old and new words also on the basis of lexical context. New words are defined as words produced for the first time in a monologue, old words are words uttered for the second time. This implies that a word has to be mentioned literally earlier in the discourse to be classified as given information. The same observation counts for the definitions used by Horne (1991a).

2.2. The relation accented/plus focus vs. not accented/minus focus

The definitions described above suggest that new information is always accented (plus focus), and that old information is never accented (minus focus). This is true, however, only for new information: listeners generally do judge it inacceptable if new

information is not accented (Nooteboom & Kruyt 1986). Old information. however, ~can in some cases be accented and thus be plus focus. This implies an asymmetry between new/plus focus and old/minus focus. Nooteboom & Kruyt (1986) explain this by assuming that there are other reasons for intonational focusing than newness. A plus focus constituent can be associated with given information depending on the word order or the constituent structure. Focusing given information can be used to highlight the theme or topic of a sentence, but only when another plus focus domain appears later in the same sentence. Horne (1991b) explains the fact that there is no strict correlation between new/plus accent and given/minus accent by means of rhythm. Given information that is focused can signal thematicity (following Nooteboom & Kruyt), but only at the beginning of a constituent, not at the end. This explanation does not cover all the cases, and according to Horne the focusing of given information is a phonological issue, that is rhythmically motivated.

2.3. Conclusions concerning focal structure related to acoustic features

The definitions described above are not sufficient to determine the structure of a whole text. The material used in the different phonetic experiments consisted of pairs of sentences or question-and-answer combinations. in which a structure of plus or minus focus is relatively easy to detect. However, if we want to analyse a discourse in terms of focus, we will need a system which is more accurate and subtle in assigning focal structure. Discourses are more complex than just a combination of sentences or question-and-answer pairs. Such a system should be able to make more distinctions than just plus or minus focus. As mentioned before, the way in which the definitions above are used present a circularity. The various experiments mentioned above had the intention to investigate the *acoustic* features of focus. However, acoustic features such as intonation and duration were already included in the definition of focus itself.

The actual experiments from the studies described above will not be presented. At this point, we are only interested in the way different notions such as focus and new vs. given information were defined in various phonetic studies.

In the next section we will present some approaches about how to present the structure of discourses, based on textual analysis rather than on acoustic measurements. The starting point is the literal transcription of a spoken text or discourse, instead of the forced focus distributions in the form of the usually used question-and-answer pairs. In this way, we should get a method that is much more able to detect the structure of spoken texts, and that takes into account more than just the focal status of some concepts or items in a sentence.

3. Theories in which the focal structure of a text is *not* determined by acoustic features

3.1. Introduction

This section will present three theories about the structure of text, based on textual analysis: 1) Rhetorical Structure Theory (RST), introduced by Mann & Thompson (1988), 2) the approach of Chafe (1987) and 3) the one used by Prince (1981). These theories were already briefly mentioned in the introduction. These approaches mainly focus on the coherence relations in a discourse, and on the ways to represent them.

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' (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 in 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. This theory accounts for the status of concepts within so called 'intonation units', these can be active, semi-active or inactive. The theory proposed by Prince (1987) is even more specific than Chafe's. This type of analysis results in a functional description that is very detailed, and is based on the 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' in Prince's system if the speaker formulates this referent as 'brand new'.

The theories described below are thus presented in an order from rather general to more specific.

3.2. Rhetorical Structure Theory (1988)

3.2.1. Introduction

First of all we present the model developed by Mann & Thompson (1988 for the definitions), Rhetorical Structure Theory. RST 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; it is not yet clear how RST can be applied to dialogues. Studies which have used RST revealed a number of advantages: relations among clauses can be described whether or not they are grammatically or lexically signalled; RST is applicable to a wide range of text types and to narrative discourse; it enables to investigate Relational Propositions, on which text coherence depends (see for instance Mann & Thompson 1992, Abelen, Redeker & Thompson 1993, Redeker 1993).

3.2.2. Description of RST

RST has four objects defined: relations, schemas, schema applications and structures.

1. Relations
Relations hold between two spans of text (non-overlapping) which are called 'nucleus' and 'satellite'. The four fields that each relation consists of are constraints on nucleus, on satellite, on the combination of both and the effect. Each field specifies judgements that the analyst must make when building the RST structure. These are judgements of plausibility.

2. 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, as pictured below in figure 1. The curves represent the relations, the straight lines identification of the nuclear spans. Other schemas all follow the pattern of a single relation with a nucleus and a satellite.
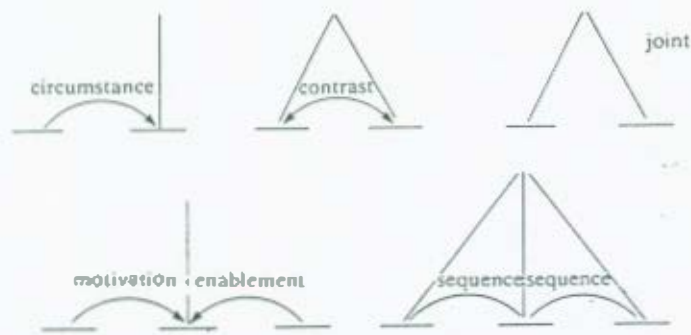
Fig. I. Examples of the five schema types (Mann & Thompson 1988, p. 247).

3. Schema applications

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).

4. Structure

A text is divided in 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, contrast. This set is considered to be open. See Mann & Thompson (1988) for examples of text analyses.

The relation definitions described above can be classified in a two-way distinction in *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' (elaboration, circumstance, solutionhood, volitional and non-volitional, purpose, condition, interpretation, evaluation, restatement, summary, sequence, contrast). 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' (motivation, antithesis, background, enablement, evidence, justify). This division is the one proposed by Mann & Thompson, others are possible as well.

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.

6

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

1. virtually every text has an RST analysis;
2. there are certain text types which characteristically do not have an RST analysis, for instance laws, contracts, poetry;
3. 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:

1. structural relations are not necessarily expressed in clauses;
2. such relational propositions can be signalled by conjunctions or other morphemes, but they can also be conveyed without;
3. the relational propositions correspond to the relations of the RST structures of the text;
4. the relational propositions are essential to the coherence of a text: if these are disturbed, the text will become incoherent.

The relational propositions are considered as being derived directly from the relation definition itself.

Mann & Thompson also present evidence for nuclearity. Earlier, the notions of nucleus and satellite were introduced. 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 analysed 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 analysed by Mann & Thompson. 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 this memories. The nucleus is the part that is most deserving of response, including attention and reaction. The nucleus is more central than the satellite.

3.2.3. Conclusions on RST

The RST turns out to be a very useful method to analyse 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 analyse a text or discourse on the level above the sentence. In our own project we will first make a rough analysis in terms of 'functional units', following the RST rules. These 'unitization rules' form a preliminary step before the RST analysis, and do not form a part of the actual analysis itself. It is however not necessary for us 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 use of the theories of Chafe and Prince to determine the structure of texts on the sentence level and below. The division of the text above the level of the sentence is needed to account for certain boundary effects, as will

become clear in section 3.5. The presence and the place of these boundaries may ~follow from the RST analysis. Therefore, we have given above a rather detailed description of Rhetorical Structure Theory.

3.3. Chafe (1987)

Chafe (1987) proposes an approach to analyse 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 best be described in this section, because the basis of the theory is the analysis of a transcribed spontaneously uttered text rather than the acoustic measurements of the speech signal.

A piece of (transcribed) spoken language naturally divides itself in *intonation units* (a single focus of a speaker's consciousness; cf. *idea unit* in Chafe 1980). An intonation unit (or idea unit) contains concepts: the ideas of objects, events and properties. Such a concept may be in one of three states at anyone 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 in 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 can not be pronominalized. Concepts marking a contrastive accent can not 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, it stays in the peripheral memory for a 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 in *paragraphs* is made through the location of responses from the hearer and through pausal evidence. *Sentences* are defined by the occurrence of 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 theories described in the phonetic experiments. The distinction used by Chafe (1987) is ternary instead of binary, and thus 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 it is assumed to be generally known.

8

This theory, in contrast with the 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 in section 3.1, Chafe's analysis concerns primarily the activation state of a concept in the head of a hearer. The analysis, however, still makes some use of acoustic features (though not as evidently as the theories described in section 2): clauses are defined as 'intonation units', which are detected by 'pauses', sentence 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.

3.4. Prince (1981)

3.4.1. Introduction

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 at three levels: in the sentence, in the discourse and in the discourse model used by the participants. At 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). These three levels are discussed, and on that basis 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 different names, for instance: given-new, old-new, known-new, presupposition-focus. However, these notions have never been characterized satisfactorily in a way to enable researchers to use them adequately and make them operational. We will present the definitions used by Chafe (1976), Clark & Haviland (1977), Halliday (1967) and Kuno (1972), 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 in her article between three types of givenness: 1. givenness as predictability/recoverability (givenness$_p$), 2. givenness as salience (givenness$_s$) and 3. givenness as shared knowledge (givenness$_k$). These types are discussed below, and the different definitions used in the literature are integrated in this tripartition.

3.4.2. Three types of givenness

1. 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 (1972) 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'. Halliday (1967) defines given-new differently, in terms of intonation: given is defined as 'the complement of a marked focus'. New information is 'information that the speaker presents as not being recoverable from the preceding context.' 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.' 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.' Prince concludes that it is crucial to consider the speaker's hypotheses about the hearer's beliefs and assumptions in the notion of givenness.

2. 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' and new as 'what the speaker assumes he is introducing into the addressee's consciousness by what he says'. This presents a binary distinction. Furthermore, a given element must have an explicit referent in the discourse.

3. Givenness$_k$ 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) defined 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.'

Kuno (1972) introduced the notions of anaphoric and non anaphoric. These also fall under the term of givenness$_k$. An element is anaphoric, if '[its] referent has been mentioned in the previous discourse' or is 'in the permanent registry' (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 of the understanding of the givenness in the sense of shared knowledge.

3.4.3. The model of 'assumed familiarity'

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).
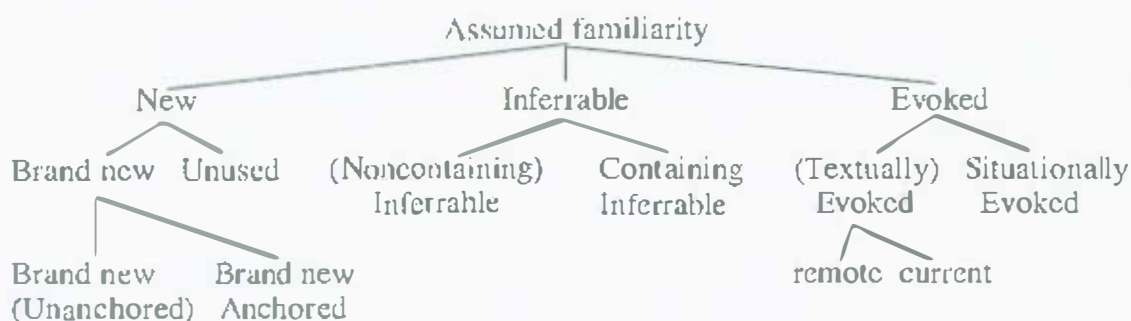
A new entity can be *brand new* (cf. to be bought in a store) or *unused* (cf. to be token from a shelf). The brand new entities can be *anchored* (linked by means of

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 NP's. This means that indefinite NP's are classified as brand new, while definite NP's are usually classified as unused and can never be classified as brand new (cf. also Vallduví 1993, p. 25).

NP's 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 inferenced 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).

The following diagram presents the different discourse entities:

Assumed familiarity

New   Inferrable   Evoked

Brand new   Unused   (Noncontaining)   Containing   (Textually)   Situationally
                     Inferrable   Inferrable   Evoked   Evoked

Brand new   Brand new   remote   current
(Unanchored)   Anchored

The textually evoked items can be further divided into *remote* and *current* (Redeker, personal communication) The remote textually evoked items are too far back in the discourse to be pronominalized (cf. semi-active in Chafe's theory), while the current textually evoked items can reoccur in the form of a pronoun (cf. active in Chafe's theory).

3.5. Comparison of the theories of Chafe (1987) and Prince (1981)

The division in three basic parts used by Prince is roughly comparable to the division used by Chafe. 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 predict 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 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 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 than the analysis proposed by Chafe to indicate focus without making use of acoustic features. The Prince analysis, however, does not apply to verbs or adverbials. In Chafe's analysis, adverbial constituents can be classified as certain 'orientations'.

4. Proposed method

4.1. Introduction

This section will present the method we intend to use to indicate focus without using acoustic features. This method contains elements from the three theories described above: RST as well as the theory of Chafe and of Prince.

4.2 Method for labeling a text

A first step in analysing a text or discourse is its division in functional pieces of text, on the basis of functional and semantic criteria. This results in a rough structure, in which major boundaries as sentence and paragraph boundaries are detectable, purely on the basis of the linguistic representation. The paragraphs are numbered, and within these paragraphs sentences are indicated by using typographical means. This will become clear in the example below.

The next step is to detect clauses which function as 'added information' or which contain comments expressed by the speaker. Also labeled are the return points: the point where the story continues after a comment by the speaker. So called 'orientations' are labeled as well. An orientation refers to an expression of time at the beginning of a clause. These labels were collected from both RST analysis (return points and paragraph; our definitions) and from the analysis proposed by Chafe (added information, orientation), and refer to the preceding clause as a whole, and to larger units.

The next step is to label all nominal elements according to the model of assumed familiarity. This results in an analysis in which every noun phrase is labeled according to the information it expresses, thus yielding a detailed analysis on the level of the sentence. The labels at this level refer to nouns plus possible determiners, not to clauses.

Prince's theory does not include any labels for elements like verbs or for adverbs. We think that these elements can express valuable information as well, and therefore we propose one more label and an extension of one of Prince's labels. Adverbs or other adverbial expressions of time or place (not sentence initial) will be labeled as 'modifier', and can contain new information depending on the context. Sentence *initial* adverbs or adverbial expressions are labeled as 'orientation'. The label 'modifier' is introduced as a new label. If verbs are used as nominalized verbs, they can easily be classified as nouns. This will be necessary only if the verb expresses new information: the information is crucial to the comprehension of the story as a coherent discourse. In that case, verbs are labeled as nouns. The label 'brand new anchored' will not be included in our method, since we did not find clear examples of this label in our texts. At this point, then, it seems not necessary to maintain this label.

The different labels to be used in our analysis are summarized below. The application of the labels will become clear in the example presented in section 4.3.2.

Label:	Function in the analysis:
1, 2, 3 etc	divides the text in functional pieces
[enter]	clauses are separated by [enter] within a functional piece
#	paragraph boundary
=	continuation of the story after interruption (ai or seg)
or	orientation (sentence initial)
ai	added information
seg	segment (comment by the speaker, metalinguistic)
bn	brand new
u	unused
i	(non-containing) inferrable
ic	containing inferrable
ct	evoked textually (current)
etd	evoked textually displaced
es	evoked situationally
mod	modifier (not sentence initial)

4.3. Pilot experiment

4.3.1. Introduction

We have conducted a pilot experiment with two versions of the same text, of equal structure, containing the same words and formulations and only differing in speaking style (spontaneous vs. read aloud).

Our first step was to make a textual analysis to determine which words or word groups were put in focus by the speaker. This analysis was done by the author according to the method described above: 1) division in major functional parts (cf. RST) and 2) within these parts different words or word groups were labeled according to the status of the information they express (cf. Chafe and Prince). Goal of this pilot study was to test the method of structure analysis, and to determine the procedure to be used in our next experiment: the evaluation of the structure of texts on the basis of spoken vs. written material. This will be done in different speech conditions, as indicated in the following scheme:

```
written:    original text                transcription of retold story
               |                        ^                     |
               v                        |                     v
spoken:     read aloud    --->        retold              read aloud
```

Subjects will be presented with the written texts with or without the three spoken discourse versions. The task is to indicate on paper the structure of the information flow in the presented text, either on the basis of the written text alone, using linguistic knowledge and intuition, or on the basis of the written text in combination with the spoken version heard over headphones. Our assumption is that in the spoken discourse versions, the linguistic intuitions may in some cases be overruled by the actual speech sound.

4.3.2. Example of a text-analysis

We used a short story by Simon Carmiggelt in Dutch ("Een triomf" from *Fluiten in het donker*, 1966). The text is analysed according to the method and approach described in section 4.2. We will present here the initial sentences from two versions of the same story: the original text and a retold version. The labels are indicated between brackets, and refer to the preceding noun with possible determiner, or to the preceding clause as a whole (ai and or).

Part of the original text:

1. Toen deze winter [u] de sneeuw [u] eens zo overvloedig [mod] begon neer te dwarrelen [or],
spoorden we [es] de stad [u] uit
om te kijken hoe het er in het bos [u] uitzag. #
2. Het was geen vergeefse reis [i]. #
3. Onder de vracht [ic] van sneeuw [et] en ijs [i] beladen [or]
kraakte het woud [et] als een orthodox spookhuis [bn].
4. Het feit [bn]
dat je [u] grote takken [bn],
die het niet langer konden volhouden, afbraken en naar beneden stortten [ai],
=op je kop [i] kon krijgen,
=gaf onze tocht [et] een accent [bn] van gevaarlijk leven [bn],
dat in de stad [et] alleen de zebrapaden [bn] kunnen bieden [ai].

Part of the retold version (by speaker 1 from Koopmans-van Beinum, 1980):

1. ik [es] heb laatst [mod] een verhaal [bn] gelezen
nou onlangs zeer onlangs [or] mag ik [es] wel zeggen [seg],
=verhaal [etd] gelezen van Carmiggelt [u] uit een bundel [bn],
een [ic] van de vele bundels [i] die hij [et] gepubliceerd heeft,
2. eh het [et] droeg de titel [u] een triomf [bn],
en het [et] was weer een typisch Carmiggelt verhaal [i] [seg],
3. eh de man [u] is namelijk in staat om allerlei eh menselijke situaties [bn] in zijn eigen woorden bn] op een bijzonder charmante en prettige manier [bn] weer te geven [ai]. #
4. eh het [et] speelde in de winter [u]
5. u [es] weet hij [et] vindt zijn stof [bn] veelal in zijn naaste omgeving [bn] [seg]
eh ... in zijn gezin [i], bij zijn kinderen [i], zijn kleinkinderen [i] [seg],
6. dit was een verhaal [etd] over hem [et] en zijn vrouw [bn]
die ergens [mod] op een uur [bn] afstand van Amsterdam [u] aan het wandelen [bn] waren in de bossen [u]
die zwaar onder de sneeuw [u] lagen [ai]
waar ze eh onder gevaarlijke omstandigheden [bn] wandelden,
tenminste als ik [es] hem [et] mag geloven [seg],
want hij [et] beschreef in allerlei lyrische bewoordingen [bn] het gevaar [u] waaraan zij [et] blootstonden [ai],
=van onder de last [bn] der sneeuw [ic] afbrekende takken [bn],

These two examples illustrate the proposed method for analysing a text in terms of focus, without using acoustic features. This analysis will constitute the starting point for our hypotheses about the relationship between the labels from this analysis and possible acoustic features. The analysis has resulted in various labels, which can be related to various acoustic features. This will lead to predictions like: 'if a clause contains a label x, it will be pronounced with y features'. These features correspond to the usual prosodic features such as fundamental frequency, duration, intensity, and spectral aspects. At this point in our project, we have not yet fully formulated the hypotheses.

5. General conclusion

5.1. Concluding remarks

In the literature focus is generally detected on the basis of intonation. This definition leads to circularity, since possible acoustic features of focus are already included in the definition itself. In this paper we propose a way to detect the focal structure of a text or a discourse that is not based on any acoustic feature. This should result in a definition of focus that is more operational for various disciplines.

The definitions presented in section 2 literally use the term 'focus'. The theories presented in section 3 do not use this term, but rather speak of the kind of information a certain element expresses. The different kinds of information can, as we see it, be linked to possible types of accentuations. In that case, accentuation is the result of the state information is in, instead of the other way around (the state of the information is the result of the accentuation, as presented in section 2). This seems to yield a more objective and in any case more operational way to approach the issue of focus.

5.2. Related research areas

This paper has not discussed any research done in the area of synthetic speech or text-to-speech systems. In these areas, however, some of the same issues are investigated as we do: how can focus be implemented in a text-to-speech system, and how can the location of accents be predicted in sentences or discourses? To be able to answer these questions, a method is needed to account for the internal focal structure of texts, without making use of acoustic features. This element is crucial, since precisely these acoustic features are subject to possible manipulation.

It would be beyond the scope of this paper to describe the various researches done in the area of speech synthesis (for instance Quené & Dirksen 1990, Hirschberg 1990, House & Youd 1990, Quené & Kager 1993, Dirksen & Quené 1993, Horne et al. 1993). We are aware of the importance of these researches, and they will be included in our own project in a later stage.

Vallduví (1993, 1994) proposes a more semantic approach to account for the 'packaging of information': "[...] it is assumed that information states are highly structured objects that allow - or even require - information to come with (un)packaging instructions" (Vallduví 1994, p. 23). The paper investigates the possible kinds of instructions that are found in communication, and suggests "a particular internal structure for information states that seems to accord with the nature of these instructions" (id. p. 23). We will not discuss these theory here, since it is not

of direct importance to this paper. We will, however, include this kind of research in a later stage of our project.

Acknowledgements

References

Abelen, E., G. Redeker & S.A. Thompson (1993) 'The rhetorical structure of US-American and Dutch fund-raising letters', *Text* 13 (3), p. 323-350.

Chafe, W.L. (1976) 'Givenoess, contrastiveness, definiteness, subjects, topics, and point of view', in: C. Li (ed.) *Subject and topic*, Academic Press, New York, p. 25-55.

Chafe, W.L. (1980) 'The deployment of consciousness in the production of a narrative', in: W.L. Chafe (ed.) *The Pear Stories: Cognitive, cultural, and linguistic aspects of narrative production*, Ablex, Norwood, N.J., p. 9-50.

Chafe, W.L. (1987) 'Cognitive constraints of information flow', in: R.S. Tomlin (ed.) *Coherence and grounding in discourse*, Typological studies in language 11, John Benjamins Publishing Company, Amsterdam/Philadelphia, p. 21-51.

Chomsky, N. (1971) 'Deep structure, surface structure, and semantic interpretation', in: D.D. Steinberg and L.A. Jakobovits (eds.) *Semantics: An interdisciplinary reader in philosophy, linguistics and psychology*, Cambridge University Press, Cambridge, p. 183-216.

Clark, H. & S. Haviland (1977) 'Comprehension and the given-new contract', in: R. Freedle (ed.) *Discourse production and comprehension*, Lawrence Erlbaum Associates, Hillsdale, N.J., p. 1-40.

Dirksen, A. & H. Quené (1993) 'Prosodic analysis: The next generation', in: V.J. van Heuven & L.C.W. Pols (eds.) *Analysis and synthesis of speech*, Speech Research 11, Mouton de Gruyter, Berlin, p. 131-146.

Eady, S.J. et al (1986) 'Acoustical characteristics of sentential focus: narrow vs. broad and single vs. dual focus environments', *Language and Speech* 29 (3), p. 233-251.

Fowler, C.A. & J. Housum (1987) 'Talkers' signaling of 'New' and 'Old' words in speech and listeners' perception and use of the distinction', *Journal of Memory and Language* 26, p. 489-504.

Halliday, M.A.K. (1967) 'Notes on transitivity and theme in English. Part 2', *Journal of Linguistics* 3, p. 199-244.

Halliday, M.A.K. & R. Hasan (1976) *Cohesion in English*, Longman, London.

Hirschberg, J. (1990) 'Using discourse context to guide pitch accent decisions in synthetic speech', *Proceedings of the ESCA Workshop on Speech Synthesis*, Autrans, France, p. 181-184.

Horne, M. (1991a) 'Phonetic correlates of the 'new/given' parameter', *Proceedings of the ICPhS 91*, Aix-en-Provence, volume 5, p. 230-233.

Horne, M. (1991b) 'Why do speakers accent 'given' information?' *Proceedings Eurospeech 91*, Genoa, volume 3, p. 1279-1282.

Horne, M. et al. (1993) 'Improving the prosody in TTS systems: Morphological and lexical-semantic methods for tracking 'new' vs. 'given' information', *Working Papers* 41, Dept of Linguistics and Phonetics, Lund, Sweden, p. 208-211.

House, J. & N. Youd (1990) 'Contextually appropriate intonation in speech synthesis', *Proceedings of the ESCA Workshop on Speech Synthesis*, Autrans, France, p. 185-188.

Koopmans-van Beinum, F.J. (1980) *Vowel contrast reduction. An acoustic and perceptual study of Dutch vowels in various speech conditions*, Doct. Diss. University of Amsterdam.

Koopmans-van Beinum, F.J. (1992a) 'The role of focus words in natural and in synthetic speech: acoustic aspects', *Speech Communication* 11, p. 439-452.

Koopmans-van Beinum, F.J. (1992b) 'Can 'level words' from one speaking style become 'peaks' when spliced into another speaking style?', *Proceedings of the ICSPL 92*, Banff, volume 2, p. 1099-1102.

Koopmans-van Beinum, F.J. (1994) *Application to obtain a grant for a facultary position as post-graduate student.* IFOTT 1994.

Kuno, S. (1972) 'Functional sentence perspective', *Linguistic Inquiry* 3. p. 269-320.

Ladd, D.R. (1980) *The structure of intonational meaning,* Indiana University Press. Bloomington.

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

Mann, W.C. & S.A. Thompson (1992) *Discourse description: Diverse linguistic analyses of a fund-raising text.* Benjamins, Amsterdam.

Nooteboom, S.G. & J.G. Kruyt (1986) 'Accents, focus distribution, and the perceived distribution of given and new information: An experiment', *Journal of the Acoustical Society of America* 82 (5), p. 1512-1524.

Prince, E.F. (1981) 'Toward a taxonomy of Given-New information', in: P. Cole (ed) *Radical Pragmatics.* Academic Press, New York, p. 223-255.

Quené, H. & A. Dirksen (1990) 'A comparison of natural, theoretical and automatically derived accentuations of Dutch texts', *Proceedings of the ESCA Workshop on Speech Synthesis,* Autrans, France, p. 137-140.

Quené, H. & R. Kager (1993) 'Prosodic sentence analysis without parsing', in: V.J. van Heuven & L.C.W. Pols (eds.) *Analysis and synthesis of speech,* Speech Research 11, Mouton de Gruyter, Berlin, p. 115-130.

Redeker, G. (1992) "Kleine woordjes' in spontaan taalgebruik - stoplapjes of signalen voor de lezer/luisteraar?" *Toegepaste Taalwetenschap inArtikelen* 43, p. 55-65.

Redeker, G. (1993) *Coherence and structure in text and discourse.* Unpublished manuscript, Tilburg University, August 1992.

Selkirk, E.O. (1984) *Phonology and Syntax: the Relation between Sound and Structure.* MIT Press, Cambridge, MA.

Terken, J. (1984) 'The distribution of pitch accents in instructions as a function of discourse structure', *Language and Speech* 27 (3), p. 269-290.

Vallduví, E. (1993) *Information packaging: A survey,* Center for Cognitive Science & Human Communication Research Center, University of Edinburgh. Draft version, June 1993.

Vallduví, E. (1994) 'The dynamics of information packaging', *Parametric variation in unification-based perspective,* DYANA-2 R1.3.B, Task 1.3, subtask 1, Center for Cognitive Science, University of Edinburgh, Sept. 1994