An evaluation of GoalGetter's accentuation

IPO
PO Box 513, 5600 MB Eindhoven

08.01.1997

Rapport no. 1142

An evaluation of GoalGetter's Accentuation

Daniël Nachtegaal

Voor akkoord: Prof.ir. S.P.J. Landsbergen

"An evaluation of GoalGetter's Accentuation"

Daniël Nachtegaal

Doctoraalscriptie Taal, Spraak en Informatica

Katholieke Universiteit Nijmegen, 1996

Preface

The research as well as most of the writing of this paper was performed at the Institute for Perception Research (IPO), in Eindhoven. I would like to thank my supervisors at the IPO, Jan Odijk and Jan Roelof de Pijper, for their support and guidance. They helped me to get acquainted with the GoalGetter system and commented on the earlier versions of this paper.
Furthermore, I would like to thank my supervisor at the University of Nijmegen, Joop Kerkhoff, as well as Toni Rietveldt and Erwin Marsi, who all were very willing to help me and comment on the proceeding of the investigation.
Of course, this research would not have been possible without the help of the speakers, Jouke, Dunja, Robert, Paula, Susan, Jeroen and Erik, nor without the cooperation of the listeners.
Finally, I thank all those who endured my presence during these seven months and who helped me through the hard times.
My special thanks and love goes to Jeanette, for being there when I needed her most.

Table of Contents

Chapter 1: Introduction

1.1 Introduction

The research described in this paper was in the area of prosody. It was directed towards the Prosody Component of the *Data-to-Speech* system *GoalGetter.* GoalGetter generates spoken reviews of soccer matches. The data for the generation of these reviews are available on "teletekst", which is a feature available on many television sets offering an extra news service in text format. A possible application could be, for example, a feature in a car stereo system. This system was ported from a Data-to-Speech system for English, called DYD (Dial Your Disc). DYD was intended to be a system that could provide users with spoken information on recordings of the work of Mozart as well as background information about the composer and his life.

The notion *Data-to-Speech* is used for speech synthesis applications that generate speech from structured data. The difference with Text-to-Speech synthesis is the fact that (semantic) information about relations between the elements of the data is available. Because texts are generated entirely, syntactic structure is known. This in contrast to Text-to-Speech systems that have to perform some kind of syntactic and semantic analysis first, which is never fully reliable.

GoalGetter's *Prosody Component*, from now on referred to as GGPC, takes care of the assignment of accents. The rules for the assignment of accents are adapted from Arthur Dirksen's *PROS-3* (1992), which is based on *Focus-Accent Theory.* Focus-Accent Theory explains the distribution of accents based on semantic-pragmatic properties of sentences and the syntactic structure of the constituents of those sentences. PROS-3 is a formalisation of the rules, stated in Focus-Accent Theory. Furthermore, a discourse model as proposed by Van Deemter (1993) is implemented to handle *deaccentuation.* Deaccentuation is the phenomenon that, under certain circumstances, words or phrases that would normally have been accented are assigned no accent. A part of Focus-Accent Theory, called *Rhythm Rule*, is not implemented in GoalGetter. This rule handles deaccentuation influenced by rhythmic properties of utterances.

1.2 Goal and questions of this investigation

The goal of this investigation was to evaluate the algorithm for the assignment of accents of GoalGetter's Prosody Component.

The domain of GoalGetter is very narrow. Only the basic information concerning soccer matches is conveyed. This is synthesized in reading style, comparable to reviews read aloud by a news reporter. It should be noted that this is different from fully spontaneous speech.

The question that needed to be answered here was: "do the prosody rules of GoalGetter handle accentuation in this domain in a satisfying manner?". The criterion to answer this question was comparison of the assignment of accents by these rules to accentuation by humans. This led to the following questions:

Q1. On which words does the accentuation by the speaker deviate from the accentuation by GoalGetter?

Q2. What are the possible explanations for these differences?

A possible second objective was to investigate whether accents can be divided in terms of 'weak' and 'strong' and whether this categorisation is based on properties of a semantic or a syntactic nature.

Q3. Is it possible to subdivide accents in 'strong' or 'weak' accents?

Q4. If the answer to Q3 is positive, what are the criteria for this division?

1.3 Method

In this investigation, I compare the perception of 'expert listeners' of accents produced by native speakers to the assignment of accents by GoalGetter. Seven native Dutch speakers were asked to read texts generated by the text-generation module of GoalGetter. Besides being native, the speakers had to be naïve with respect to the field of prosodics and phonetics. This prevented speakers from applying theoretical rules instead of unconscious, natural rules.

The recorded utterances were presented to nine 'expert listeners', people acquainted with prosody and phonetics, who were asked to indicate on which words they heard accents. The use of expert listeners was preferable in this case because they did not need a lot of instruction and training to produce useful data. This limited the noise in the data without taking too much time. The listeners were asked to indicate wether they heard a 'strong' or a 'weak' accent.

The texts had to be selected from an amount of generated texts. On the one hand, this selection was necessary because only a limited amount of speech material could be presented to the listeners due to time limitations. On the other hand, a selection was required because flaws in the text generation module would otherwise cause the speech material to be less useful.

1.4 Results

The results of this investigation are as follows: the number of words on which the accentuation by the speaker deviated from the accentuation by GoalGetter is very small. This shows that the agreement between the assignment of accents by GoalGetter and that by the speaker is high. The differences found will be explained in chapter 6 and solutions or suggestions to improve GGPC will be given. Furthermore they show that it is not possible to make a distinction between weak and strong accents with the experiment performed. This makes the last question of section 1.2 irrelevant.

1.5 Relevance of the results

As mentioned in section 1.3, the texts that were used in this investigation were selected to obtain useful speech material. This selection and the narrowness of the domain of GoalGetter had as a consequence that not all linguistic constructions which can be constructed by GoalGetter, occur in the data. So, not all varieties of accentuation occur in the speech material. Furthermore, accentuation by the speakers depends heavily on the quality of the text generation, because it affected the naturalness of the texts that had to be read aloud. Unnatural text structure will lead to accentuation that will not be encountered in natural speech. Therefore, the results will only be valid for this application. Furthermore, the applicability for this application of the results will be limited because not all possible constructions occur in the data.

The material on which the experiments were performed is representative for the domain of GoalGetter to a certain extent. It should be kept in mind that, within the borders of naturalness, a speaker has some freedom in producing accents. Therefore one speaker was selected, who, according to most of the listeners, had the best way of reading the texts aloud. The results can and probably will be different when other speakers will be used.

1.6 Structure of this paper

The next chapters will report on this investigation. Chapter 2 is an introduction in which the concepts of *accent* and *focus* as used here are explained and *Focus-Accent Theory* is summarized. In chapter 3, GoalGetter will be explained in detail. Its architecture will be described and the way the speech is derived from the input data. Chapter 4 describes the way in which the natural speech data were collected. The processing of the data and the results are given in chapter 5. These results are discussed in chapter 6. There, an attempt will be made to explain the results and give suggestions for improvement. Finally, in chapter 7 general conclusions will be given as well as recommendations for further investigation in this direction.

Chapter 2: Theoretical background

In this chapter I will introduce the notions of *accent* and *focus*, and I will go into *Focus-Accent Theory*.

2.1 Prosody

Terken and Collier (1995) define prosody as "... the ensemble of sound attributes that do not constitute the phonetic identity of individual speech segments, but rather encompass larger units such as syllables, words, sentences and even paragraphs." Aspects of prosody are the variations in pitch, duration, loudness and timbre. These aspects combine to produce the sound attributes mentioned above, which are speech melody, also called intonation, accentuation and emphasis, tempo and rhythm, and voice quality.

The prosodic aspects can be distinctive. Terken and Collier (1995) give an example of this fact with the word sequence 'John says Peter is a liar'. This can be pronounced in such a way that it means: 'John says: Peter is a liar.' or 'John, says Peter, is a liar'. Therefore, good prosody is very important to the correct comprehension of utterances. In addition, poor prosody hampers the listeners in their understanding of the utterances or even prevents them from understanding the utterances at all. Therefore, in synthesizing speech, the prosodic properties of the text have to be known. These are determined by syntactic, semantic, contextual and metrical properties. The latter properties of the text therefore have to be available to obtain prosodically adequate spoken texts.

Two important ingredients of the prosodic structure are phrasing and prominence cues. These make the syntactic and semantic structure more transparent to the listener by providing explicit cues to the syntactic structure of the utterance (cf. the example presented above).

Prosodic phrasing refers to the way in which sentences and texts are segmented into smaller units. Though phrasing is related to the syntactic properties of the text, prosodic constituents are not mapped one-to-one on syntactic constituents. This causes problems when locating the prosodic boundaries. Several theories have been developed, for instance by E. Selkirk (1986) or Nespor and Vogel (1986), that counter this problem. But, as phrasing is outside the scope of this paper, I will only refer to these authors without discussing their theories.

2.2 Prosodic prominence and accents

Prosodic prominence refers to the strength relation between elements within a given domain. When two elements are at the same level, one of them is prosodically more prominent than the other. This is usually expressed in terms of weak/strong relations. These prominence relations are relevant for the assignment of speech rhythm and accentuation.

Prosodic prominence is used by the speaker to convey syntactic and semantic structure of his message to the listener. It is marked by the presence or absence of pitch accents.

2.2.1 Focus

Focus is a semantic-pragmatic property of word groups. For a proper synthesis of speech it has to be known which constituents are in focus. Therefore, some kind of semantic structure has to be known. Constituents can be either in focus or not, and are labelled [+F] or [-F] respectively to mark this.

A speaker puts those constituents in focus that he thinks contain important information and should be highlighted. When a sentence is introduced in a context, it contains both new information and information that connects to the preceding information. Usually, the new information will be in focus, the given information will not. By leaving elements unfocussed a speaker assumes that the listener is familiar with the information or should be able to derive it from the preceding context. However, a speaker can decide to leave a new piece of information unfocussed in order to focus upon a certain contrast between two other elements. By contrasting certain elements in a sentence, the focus shifts to these elements.

2.2.2 Accent

Accent is a property of words (or better: of syllables). Accents are assigned at the phonological-phonetic level. For the correct prosodic realisation of an utterance it has to be known which words accents are assigned to. Words are either accented or unaccented.

An accent is realized by means of a conspicuous pitch change, longer duration and greater amplitude, with pitch change and duration lengthening as the most important properties. Baart (1987) defines accent as: "syllable prominence caused by a conspicuous pitch rise or fall". Baart therefore speaks of 'pitch accent' as opposed to accents caused by other sound cues such as increased loudness or lengthening. (Baart, 1987, section 1.3, page 4). In the GGPC, the notion *accent* is only used for the prominence-lending pitch movement on syllables.

2.2.3 Relation between focus and accent

A speaker produces an accent as a cue to focus; an accent on a word indicates that the word group to which it belongs is in focus. Looking from the reverse perspective, if a word group is in focus, the prominence relations of its constituents decide which word is assigned an accent.

Terken and Nooteboom (1987) found that incorrect accenting and deaccenting results in slower comprehension of the utterances. The effect of incorrect deaccenting is bigger than that of incorrect accenting. They found that words expressing new information are comprehended faster when they are accented (thus in focus), words expressing given information are comprehended faster if they are unaccented. Terken & Nooteboom (1987) explain this by saying that if a subject hears a deaccented expression he will try to map it on the (restricted) set of activated referents. Referents are activated when they are mentioned before. Because the set is small, only the most recent referents are contained, the referent can be identified quickly. In case of an accented expression the set of possible interpretations is in principle unrestricted. The intended interpretation must be constructed from information contained in the speech signal, so inappropriate accentuation leads to slower comprehension. Terken & Nooteboom (1987) suspect that this indicates that people process accented and unaccented words in different ways.

2.3 Focus-Accent Theory

For the assignment of accents several approaches have been proposed. Baart (1987) summarizes the basic assumptions of these approaches. A first approach is the syntactic approach. It is assumed that there is a distinction between *normal* and *contrastive* accentuation. The normal accentuation is derived from lexical and syntactic properties of a utterance, while contrastive accentuation modifies the semantic tendency. To this approach belong, among others, Chomsky & Halle (1968), and Lakoff (1972). A second approach claims that there is no predictable relation between syntactic structure and the distribution of accents, but that the distribution of accents only depends on pragmatic or semantic principles. This is the semantic-pragmatic approach. This approach is advocated by among others Schmerling (1976) and Halliday (1967). Thirdly, there is an intermediate approach that combines both foregoing approaches. Based on the intention of the speaker, focus is assigned to constituents of the utterances. Within these constituents the distribution of accents is determined by the prominence relations derived from the syntactic structure of these constituents. This is the focus approach, which is advocated by among others Ladd (1980, 1983) and Gussenhoven (1983). Baart (1987) adopts a lot of the principles of this latter view. He explains the distribution of accents for a given focus distribution and calls the framework she describes *Focus-Accent Theory.*

2.3.1 Assignment of accents

Focus-Accent Theory states that once it has been determined which constituents are in focus, the assignment of accents is predictable from syntactic information. The distribution of accents can be accounted for by a metrical representation of the utterance. This metrical tree, also called prosodic tree, is derived from the syntactic surface structure of a sentence by copying its tree structure. The top node of a constituent is labelled for focus. All other nodes are unspecified with respect to focus. As prosodic trees are strictly binary branching, non-branching nodes as well as nodes that dominate phonetically empty material are pruned. All nodes are labelled *w* or *s* ('weak' or 'strong'). This labelling is determined by the grammatical relation between the elements. Focus-Accent Theory follows Categorial Grammar in the distinction between arguments and predicates. Arguments are strong relative to predicates.

A terminal node, which is dominated by a [+F]-node directly or along a path of only strong nodes, is called a prosodic head. Now the **Pitch Accent Assignment Rule** states:

```
(1) in a sentence, assign a pitch accent to every prosodic head.
(Baart, 1987)
```

In PROS-3, the Pitch Accent Assignment Rule is implemented recursively as:

```
Accent Rule:
   For each node X, X is accented if
   a. X is marked +F, or
   b. X is strong, and the node immediately dominating X is accented.
```

Strong nodes inherit accent from their parent node, while weak nodes are only accented if they are marked [+F].

However, the labelling of the nodes of the metrical tree has to be determined before this rule applies. Given/newness of information and contrast influence the distribution of [+F] and [-F], while rhythmic effects affect the w/s-labelling.

2.4 Deaccenting

As was mentioned in section 2.2.3 words or word groups that contain new information, and therefore are in focus, should be accented, constituents that contain given information, and therefore are unfocused, should be deaccented. Deaccenting is indicated by assigning [-F] to the nodes of the deaccented constituents.

Van Deemter (1993) describes a formalism for deaccentuation. Focus-Accent Theory does not contain such a formalism, but rather starts from an assumed focus distribution.

Van Deemter (1993) has summed up the conditions under which words should be deaccented.

Deaccent words that are:
a. *lexically unaccentable*: words that are inherently unaccentable, such as certain function words.
b. *object-given*: A phrase is given information due to an occurrence of another phrase with which the first phrase stands in a relation of Identity Anaphora.

E1. "Have you met John? John is a good friend of mine."
In this example the second appearance of John is deaccented.

c. *concept-given*: Van Deemter (1993) defines this as: "an occurrence **w** of a word is concept-given if the same or the previous sentence contains, to the left of **w**, another occurrence **w'**, of an expression, whose reference is known to be subsumed by that of **w**."

E2. "Juan owns a bicycle. You need a vehicle if you work at Stanford."
In this example, the reference of "bicycle" is subsumed by that of "vehicle". "vehicle" is therefore deaccented.

When constituents are in contrast (with something mentioned earlier) focus may shift to these constituents and the words that should otherwise be accented are now deaccented. Van Deemter (1993) marks those words with [-C]. In practice, they can be labelled [-F] too, because the result will be the same (deaccenting).

After marking of de-accentuation the Default Accent rule applies to the metrical tree. This rule turns w/s-labelling around in the metrical tree.

`Default Accent: If the strong sister is deaccented, that item is weakened with respect to its sister. Therefore this sister will be strengthened. The weak/strong-labelling is reversed. If none of the sisters can be accented, both nodes will be labelled weak.`

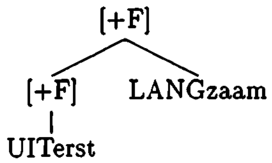
2.5 The Rhythm Rule

The w/s-labelling of trees can be influenced by rhythmic factors. Sequences of closely spaced accents cause a so called 'pitch accent clash'. The following examples, copied from Baart (1987, 104) show the existence of a rhythmical device.

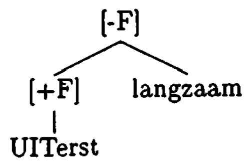
3.72 de wagen reed UITerst LANGzaam
3.73 de wagen reed UITerst langzaam

The phrase "uiterst langzaam" is labelled [w,s] and in 3.72 both the whole phrase and "uiterst" are in focus (and labelled [+F]). In 3.73 "langzaam" can only be unaccented if "uiterst langzaam" is out of focus, while "uiterst" is still in focus.

3.72

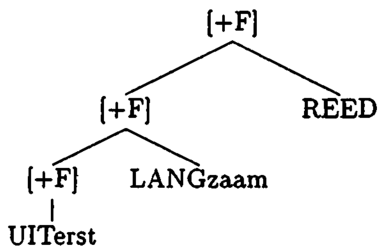


3.73

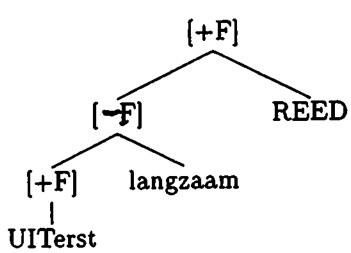


3.74 omdat de wagen UITerst langzaam REED

In this sentence "langzaam" is still part of the focused information. This would result in an accentuation of the utterance: "UITerst LANGzaam REED".

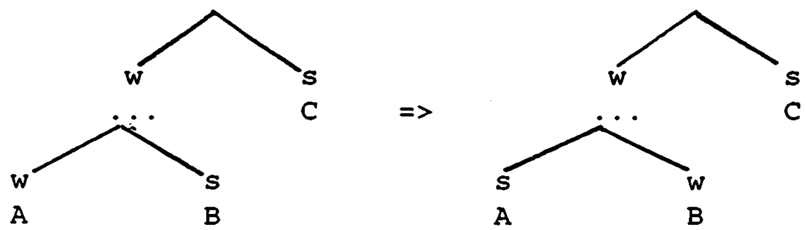


The way this phrase should be produced however is: "UITerst langzaam REED". So here we have an element left unaccented without becoming [-F]. The weak/strong-labelling of "uiterst langzaam" must have shifted to prevent an accent from landing on "langzaam".



Baart (1987) defines a rule called the Rhythm Rule. This rule was first formulated by Kiparsky (1966). The Rhythm Rule accounts for a change in the order of the w/s-labelling ([w,s] to [s,w]), when the nodes are dominated by a weak node which is followed by its accented strong sister.

Rhythm Rule (applies optionally and from left to right in a prosodic tree):



Conditions:
i. it is not the case that B is dominated by a focus while C is not and,
ii. B and C are string-adjacent, and
iii. A is not a pronoun, article, preposition or conjunction.

Baart (1987) also arguments that the Rhythm Rule must apply to hierarchical structure rather than to a linear sequence of accents. The linear formulation would state that if three or more accents are encountered in one syntactic unit the middle ones can be left out. This would not affect focus structure.

2.6 Conclusion

Focus-Accent theory provides a good means for describing the distribution of accents. The theory in itself however is incomplete, it does not account for the assignment of focus. Van Deemter (1993) describes how discourse modelling could add to this theory, by providing a systematic approach to deaccenting phenomena. Rhythmic influences on accentuation are accounted for by the Rhythm Rule as described in Baart (1987. section 3.2.3). In a diagram a complete system for the assignment of accents can be described as:

1. a metrical tree is constructed based on the syntactic tree.
2. nodes are labelled [-F] or [+F]
3. de-accentuation is marked (by assigning [-F]), Default Accent rule applied, changing the w/s-labelling of the metric tree.
4. rhythmic effects are accounted for by the Rhythm Rule, changing the w/s labelling of the nodes.
5. accents are assigned by the Pitch Accent Assignment Rule.

It should be mentioned here, that in GoalGetter, basically, this system is followed, but that it is implemented in a different way. This will be explained in the next chapter.

Chapter 3: GoalGetter

3.1 Introduction

"GoalGetter" is a Data-to-Speech system. It generates spoken reviews of soccer matches based on the information about these matches on "teletekst". The "teletekst" data is transmitted along with the regular television signals. GoalGetter puts the data found on certain "teletekst" pages about soccer into a data-structure. Texts are generated based on this structure. The individual sentences of the texts are generated by means of *syntactic templates*, from now on called, emphatic, templates. These templates define the syntactic surface structure of a main sentence, represented as a tree; these contain variable *slots*. The slots are filled with smaller trees. Such a tree is created by a function, which takes the piece of information to be expressed as its argument. The function defines another template to express that piece of information.

Before a text is passed on to the speech output module, the correct prosodic structure is attached to these sentences using linguistic and semantic information, obtained during construction of the texts. GoalGetter has two options of making the generated texts audible; diphone-synthesis and phrase-concatenation. The latter option is the concatenation of whole phrases which are taken from prerecorded speech and vary in length from a single word to several phrases. The first option is synthesis by means of the diphone speech generator "Spengi", developed at the Institute for Perception Research, IPO.

The final result should be a fluently spoken, coherent, varied and entertaining review on a game of soccer, comparable to a radio broadcast review, in which all the information that can be found on the videotext-page should be represented correctly.

3.2 The architecture

Figure 3 is a graphical representation of the architecture of GoalGetter. The boxes with round angles indicate the processes of the system.

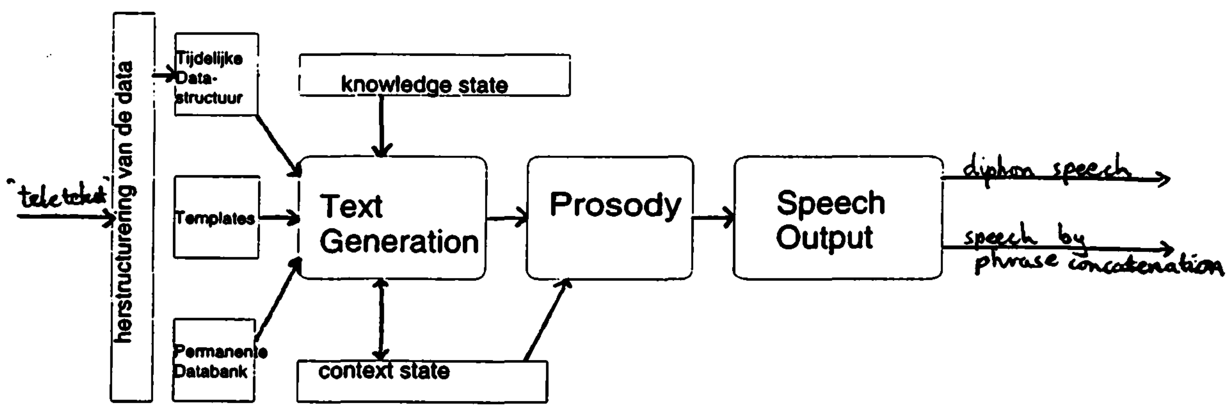


Figure 3: GoalGetter's architecture.

As is made visible in the pictorial representation above, 'Text Generation' starts from three different sources of input: a collection of templates, a database containing permanent information on teams etc. and a data structure containing information obtained from a "teletekst" page. Besides, it maintains and uses a context state and knowledge state. This module is described in section 3.4.

The 'Prosody'-module converts the syntactic trees to metrical trees. Here, it is determined where phonological phrase boundaries and accents are located. This module consults the *context state.* The result is a sequence of prosodically annotated words, called 'enriched text'. Section 3.5 describes its functioning. The 'Speech'-module uses the enriched text to produce diphone speech output or speech by *phrase concatenation.* The speech output module will be discussed in section 3.6.

3.3 The input

Input to GoalGetter is the information on soccer matches on "teletekst". One match per time is processed. Figure 1 is an example of a teletekst-page containing such information.

```
<<<<<<< Teletekst pagina: 665 01 >>>>>>>
 (C) IST NOS-TT  665 ma 18 mrt  10.18:55
   <,,(,,4,,1(,,4,1,$      VOETBAL
   -,1h,,%0 j(,,1 j PTT-TELECOMPETITIE
   ,,.*   -,.   % *

 AJAX               4 FEYENOORD        1
 Bogarde (9)          Van Gastel (41)
 Overmars (21)
 Reuser (83/pen)
 Kluivert (84)

 Arbiter:             Toeschouwers:
 Van Dijk             2.500

 Geel:
 Bogarde              Blinker
                      Vos
                      Fraser

        uitslagen 661 / stand 662
```

Figure 1: example of "teletekst" input.

The information on the "teletekst" page is presented in tabular format. The layout of the page contributes to the meaning of the information units. For example, on line eight one can get the information Bogarde (9), which means that Bogarde scored a goal for Ajax in the ninth minute of the game, and also Van Gastel (41), which means that Van Gastel scored a goal for Feyenoord in the 41st minute. This representation is transferred to a linear representation in which it does not matter what the layout looks like. Figure 2 is the translated version of figure 1.

```
team 1: AJAX;
goals 1:  4;
team 2: FEYENOORD;
goals 2:  1;
goal 1: Bogarde (9);
goal 1: Overmars (21);
goal 2: Van Gastel (41);
goal 1: Reuser (83/pen);
goal 1: Kluivert(84);
arbiter: Van Dijk;
toeschouwers: 2.500;
geel 1: Bogarde;
geel 2: Blinker;
geel 2: Vos;
geel 2: Fraser;
```

Figure 2: translation of Figure 1.

This translated file is used to create the temporary data-structure used for the generation of the texts.

Next to this temporary data representation, GoalGetter uses a permanent database and a collection of templates to generate texts. Information about players and teams, e.g. the position of a player or the name of the stadium where a team is based, is stored in the permanent database . The templates are the blueprints for the sentences of the texts.

3.4 The generation of text

3.4.1 Templates

Sentences are derived from templates. These templates define a syntactic surface structure of a sentence. An example of a template is given in Figure 3.

```
TEMPLATE Sent10
/* <scheids> deelde <aantal> gele kaarten uit */
<

LOCAUX scheids FROM express:[sum, sum.scheidsrechter, c,{},asap] / ^bearsnom

LOCAL lijst : {cardevent} = getvalue:[sum.gelekaarten] AS List(cardevent)
      hoeveelheid : Int = #lijst
      aantal : stree = mkgetaltree:[hoeveelheid, default]
LOC ks  : CODE = markexplicit:[sum.gelekaarten, c.loc],
    nc  : context = c

TREE cp[LOCAUX scheids
      cb [c0 <deelde>
          ip [np <>
            ib [vp [vb [np [LOCAUX aantal
                        nb [ap [ab [a0 <gele>]]
                            nb [n0 <kaarten>]
```

```
                                           ]
                                        ]
                                    v0 <uit>
                                 ]
                              ]
                          i0 <>
                          ]
                     ]
                  ]
             ]

TOPIC conclusion
TELLSABOUT sum.gelekaarten
CONDITION notknown:[sum.gelekaarten] AND
        known:[sum.scheidsrechter] AND
          #(sum.gelekaarten.value) >= 4 AND
        NOT Any:[known, sum.gelekaarten.value]
>
```

Figure 3 : example of a template.

This template expresses a sentence like:

(1) Scheidsrechter Van Dijk deelde vier gele kaarten uit.

The definition of a template has to meet some requirements. "TEMPLATE Sent10" tells us that we are dealing with a definition of a template, which is numbered sentence 10. The definition of this template starts with '<' and ends with '>'. The basic part of a template is called "TREE", which defines the structure of the tree derived from the template. Figure 4 is a tree representation of figure 3.

The terminal nodes of the tree part of the template can be represented as follows,

(T1) <scheids> deelde <aantal> gele kaarten uit

<scheids> and <aantal> are represented between angled brackets to indicate that they are variable slots. From now on we will specify templates in this manner. In the template of figure 3, this representation is added as comment (between '/*' and '*/'). Note that in the definition of the tree, angled brackets are used to indicate the invariable terminal nodes.

I will explain the filling of the slots in section 3.4.3. There, "LOCAUX", "LOCAL" and "LOC" will be discussed. "TOPIC", "TELLSABOUT" and "CONDITION" will be discussed in section 3.4.2.

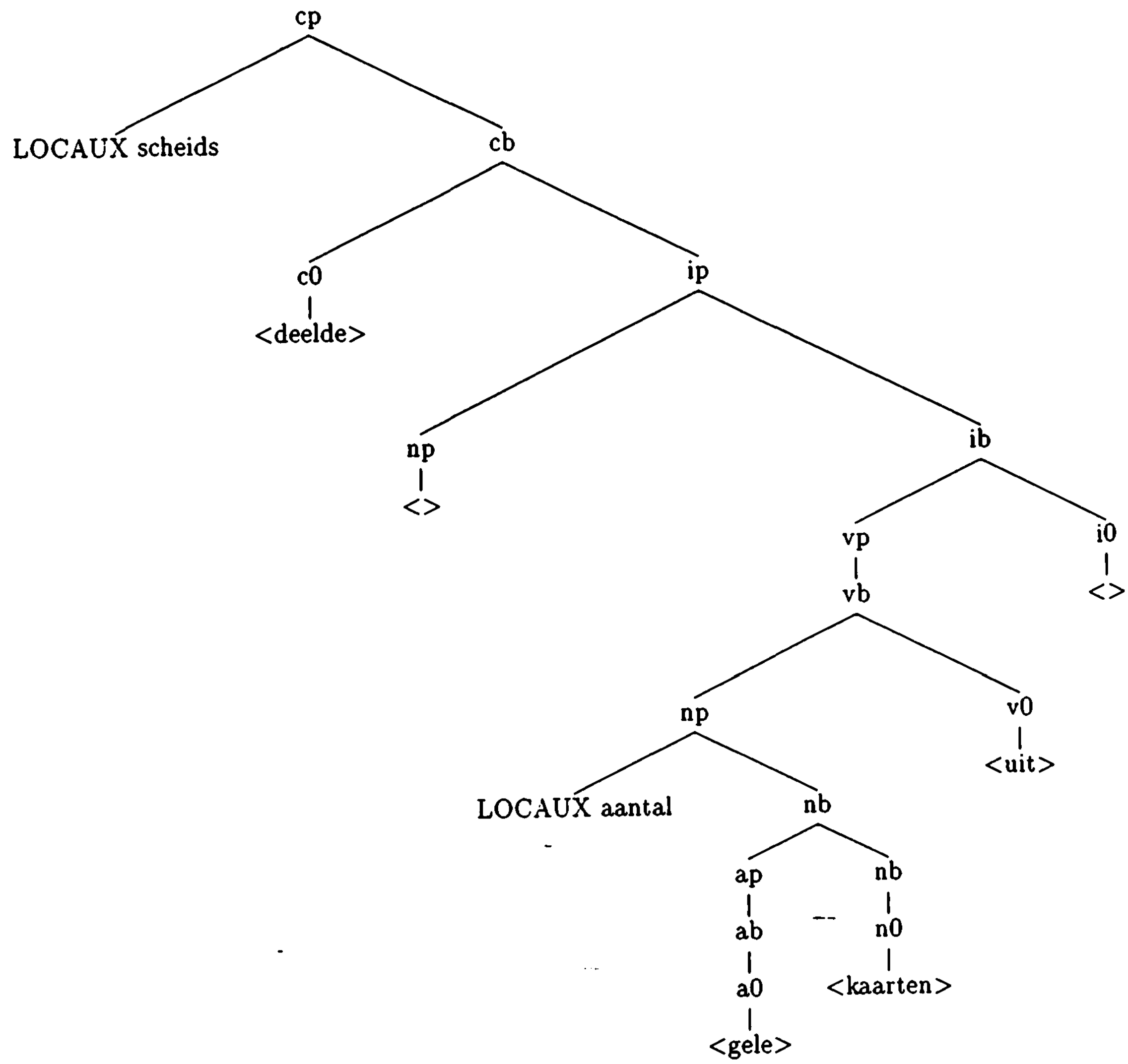


Figure 4: tree representation of the template from figure 3.

3.4.2 Conditions on templates

Before a template yields a syntactic tree, some requirements have to be met. For this purpose TOPIC, TELLSABOUT and CONDITION are added to the template.

In GoalGetter, TOPIC divides the templates into three classes, according to which paragraph of the text they should be used in. These *topics* are: "introduction", "body" and "conclusion". "introduction" is a brief summary at the beginning of the text. "body" is the temporal report of the match, summing up the events. "conclusion" is the part where some additional information is given. The template of figure 3 belongs to "conclusion". For each topic, GoalGetter searches the data representation trying to express the pieces of information using the templates appointed to that topic.

TELLSABOUT specifies a subset of the data structure about which the template tells us something. If this set is contained in a variable called "whattotell" which is used to keep track of what pieces of the information have to be told, the template can be used. In the example of template

"Sent10", TELLSABOUT specifies that the template will tell us something about yellow cards ("sum.gelekaarten"). If the variable "whattotell" contains "sum.gelekaarten", then the template can be used.

Thirdly, CONDITION states conditions on the use of templates, based on what information still has to be told or what information is already known. Which pieces of information are known is maintained by the knowledge state, see section 3.4.4. In figure 3, it is stated that nothing should be known about yellow cards ("notknown:[sum.gelekaarten]"), but it should be known who is the referee ("known:[sum.scheidsrechter]") and the number of yellow cards should be equal to or bigger than 4 ("#(sum.gelekaarten.value)>=4") and nothing should be known about individual yellow cards ("NOT Any:[known, sum.gelekaarten.value]"). "AND" is a logical operator, combining several conditions into one.

3.4.3 Slots

The slots are filled with syntactic trees derived from other templates. Functions which use these templates to form the (usually smaller) phrases are called from the basic templates. In the tree of figure 3, LOCAUX marks locations of variable slots. For example, "LOCAUX scheids" is used to indicate where the variable slot <scheids> should be inserted in the tree; "LOCAUX scheids FROM express:[sum, sum.scheidsrechter, c, {}, asap] / ^bearsnom", calls the function "express" with the list of arguments specified between square brackets. The function returns a set of triples <tree, ks, c>. "tree" is the syntactic surface structure tree, an S-tree, which is the expression that refers to the entity <scheids>. "/^bearsnom" selects the topics of which the case of the topnode is a nominative. "ks" and "c" are the changes to the knowledge state and the context state respectively after the function and the complete template is belongs to are applied.

In the "LOCAL" section the slot "aantal" is filled by the function "mkgetaltree". "aantal" is defined as an "stree". But first the argument "hoeveelheid" of type integer ("Int") is declared as the number of events of variable "lijst", which in turn is declared as a list of "cardevents" which is filled with the list of the yellow cards.

When it is not possible to fill all the slots in a template, it does not yield a syntactic tree, hence no sentence.

The variable slots introduce a source of variation. A second way of filling the template of figure 3 is, for example:

(2) Van Dijk deelde vier gele kaarten uit.

The filling of the slots has to be done conscientiously to avoid generation of incorrect or overcomplete sentences. For example, the template (the square brackets "[","]" indicate optionality, the "I" indicates choice):

(T2) <tijdsbep> scoorde <speler> [opnieuw I alweer] [voor <ploeg>],

could generate sentences like:

(3) in de 43e minuut scoorde Kluivert van Ajax voor Feyenoord.
(4) in de 43e minuut scoorde Kluivert van Ajax voor Ajax.

Sentence (3) is incorrect because Kluivert is a player of Ajax and can not score a goal for Feyenoord. The possibility of an own-goal is covered by an extra template. Sentence (4) is overcomplete, it is sufficient to specify the team just once.

3.4.4 Context state and knowledge state

A requirement is that information is not presented as new more than once. This holds both inside templates and across template borders. For this purpose a *context state* and a *knowledge state* are maintained. In the template language, the knowledge state, 'ks', and context state, 'nc' (new context), are maintained in the section called "LOC".

The knowledge state keeps track of which information already has been expressed and when. In figure 3, a new element "sum.gelekaarten" is added to the knowledge state by the function "markexplicit". (This function will be explained in more detail in section 3.4.5.) As was mentioned in section 3.4.2, it is checked in "CONDITION" whether the conditions are true with respect to the knowledge state, before a template can be used. If "sum.gelekaarten" already is contained in the knowledge state (hence it will be 'known'), the template can not be applied.

Another example of the usefulness of the knowledge state is demonstrated by the following sentence, which could be derived from (T2),

(5) in de 43e minuut scoorde Kluivert van Ajax opnieuw.

In sentence (5) "opnieuw" can only be used when Kluivert has scored the preceding goal and this has been conveyed. This information can be found in the knowledge state.

The context state records which objects have been used in the text, and how and when they have been referred to. This makes it possible to formulate rules for the proper use of referring expressions. The context state also contains all previous words in the sentence as well as all words of the preceding sentence. From the context state it can be derived which information is object- or concept-given. See section 2.4 for a definition of these notions.

3.4.5 Direct vs. indirect information representation

There are two ways to convey information present in the database. One way is to simply substitute an expression for one piece of information for a variable. The other is by computing new information from the given data. For example, it is possible to express two consecutive goals both by expressing the time of each event. Another option is to compute the moment of the second goal relative to the first.

a. in de 41e minuut scoorde Van Gastel voor Feyenoord.
b. in de 60e minuut scoorde de aanvaller Vos voor Feyenoord.

alternative to the second sentence:

b'. 19 minutes later scoorde de aanvaller Vos voor Feyenoord.

So, both "in de 60e minuut" and "19 minuten later" are possible substitutes for the slot <tijdsbep>. In the first case, time is expressed by a prepositional phrase, in the second by an adverbial phrase.

The information about the time of the goals expressed in sentences a. and b. is represented explicitly. Therefore the function "markexplicit" will be used to add it to the knowledge state. However, the time of scoring conveyed in sentence b'. is expressed implicitly by referring to the previous goal. Therefore, the actual goal should be marked implicitly in the knowledge state. This is done by the function "markimplicit". When the next goal is scored the expression of the time can not refer to this goal. Rather it should be stated explicitly or it should refer to the time of scoring of sentence a.

c. in de 63e minuut scoorde Kluivert voor Ajax.
c'. 22 minuten later scoorde Kluivert voor Ajax.

3.5 Adding prosodic cues

As stated in section 2.1, the prosodic properties of a text are determined by syntactic, semantic, contextual and metrical properties. Fortunately, in GoalGetter the syntactic, semantic and contextual properties are available, because of the generation by templates, and metrical properties can be computed.

The syntactic trees obtained by the generation process are input to the Prosody component. In this component accents and phonological phrase boundaries are computed and added to the sentences to obtain the prosodic annotations necessary for correct speech output. In order to arrive at the prosodically annotated representation, it first has to be determined which constituents are in focus and which are not. Metrical trees, M-trees, have to be constructed from S-trees. Accents have to be distributed along the M-trees and the positions and strength of the prosodic boundaries have to be determined.

As mentioned in the introduction, the accentuation of the Prosody module of GoalGetter is based on a program of Arthur Dirksen (1992) called PROS-3, which he derived from Focus-Accent Theory. However, neither the Rhythm Rule nor any other device to account for rhythmic effects on accentuation is implemented. Both Focus-Accent Theory and Rhythm Rule were explained in chapter 2. How they are implemented in GoalGetter is explained in section 3.5.2. Furthermore, GoalGetter is augmented with a discourse model, described by Van Deemter (1993), to account for de-accenting phenomena, see section 2.4.

3.5.1 Adding the focus attribute

For each node of the S-tree the value of the focus attribute is determined. By default, all nodes receive the value neutral, except for nodes that were marked to be in focus, [+F], or out of focus, [-F], in the templates. Also, all maximal projections are marked [+F].

Besides this first labelling of focus, it has to be checked whether there are words that have to be deaccented. The value of the focus attribute is changed to [-F] if (1) a word is concept-given, (2), for S-trees, the topnode of that tree is an identity anaphor (object-given), or (3) a word is lexically unaccentable.

As mentioned before in section 3.4.4, concept-givenness is determined from the context state. For each word in a sentence it is checked whether it is a member of the set of given words, or if it is a synonym of a member of this set, or if it subsumes a word present in this set. If so, it is marked as [-F].

However, some words, such as cardinals, should never be deaccented. In GoalGetter the expression of the result of a match is an example of this. In the sentence: "Ajax tegen Feyenoord eindigde in 1-1", both "4" and "1" should be accented. The general rules would cause the second "1" of the phrase "1 - 1" to be unaccented. In addition, in this particular case, the cardinals have to be realised in a specific way. For these reasons they are marked with an "@".

3.5.2 Metrical trees

Metrical trees, M-trees, are constructed from S-trees and at the same time accents are assigned and distributed. A node of an M-tree can only be unary or binary branching. If a node of an S-tree has more than two branches a dummy node is inserted which has the leftmost sibling as its sister and the other nodes as its siblings.

Each node is labelled w, 'weak' or s, 'strong'. The weak/strong labelling defines the relative prominence that two sister nodes have with respect to one another. The node labelled 'strong' is more prominent than its sister in Dutch. Almost in all cases the right sibling of a node gets the label 'strong', except when it is a zero-projection and the left sibling is not. When the node is unary branching, the sibling is labelled 'strong'.

Each node has an attribute for accent. This attribute can have the values [at], accented, [af], deaccented, [an], neutral accent or [ac], contrast accented.

An S-tree is traversed from top to bottom, for each node the syntactic category is copied to the metrical tree. If that node is labelled [+F] an accent [at] is launched, if the node has the value TRUE for an extra attribute called CONTRAST an accent [ac] is launched. This accent is distributed to the strong son, unless the right son cannot be focussed but the left son can. If the node is labelled as [-F], the accent [af] is launched, which means not accented, or deaccented. If focus is neutral or if the node is labelled [-F] and accent is [ac], contrast accent, the value of the accent is merely passed on to the strong son.

3.5.3 M-words

Words of the S-trees are converted to M-words. M-words have additional attributes to represent the relative prominence of the M-word, *prom*, and the relative strength of the prosodic boundary following the M-word, *boundary*. The value of *prom* is determined by the accent attribute.

For each M-word which immediately precedes a comma, the value of *boundary* is set to 0.5 (weak prosodic boundary). For each M-word that precedes a right clause or a punctuation symbol other than a comma, the value is set to 1.0 (strong prosodic boundary). All other M-words get the value 0.0 (no prosodic boundary). There are some other rules, but as this paper is not on prosodic boundaries, I will not give an exhaustive enumeration.

3.5.4 The output to the speech module

In DYD, for each sentence a sequence of triples is passed on to the speech output module, <str, accent, prosodic boundary>. 'str' is a string attribute containing the string of the word, 'accent' is the accent attribute with the values 0 (not accented) or 1 (accented) and 'prosodic boundary', with the values 0 , 0.5 and 1.0, giving the strength of the prosodic boundaries.

In GoalGetter, however the output to the speech module consists of enriched text. This is plain text annotated with special symbols. An accent on a word is indicated by an " before that word. Prosodic phrase boundaries are marked with /, //, or /// corresponding respectively to 'weak', 'strong' and 'end of sentence'. @ is used to signal special use of cardinals, as explained in section 3.5.3. % is used to indicate that the word followed by an apposition should be accented differently.

3.6 Speech production

The Speech module of GoalGetter contains two different options for speech output. One is synthesis by means of diphones, the other is concatenation of pre-recorded speech segments varying in length from a word to several phrases.

In the speech module the result of the prosodic module can be translated to two representations, dependent on the specified form of output. Which form of speech output is required can be specified when starting the generation process for a soccer match.

3.6.1 Diphone synthesis

For diphone synthesis the enriched text is transformed to an annotated phonetic representation. The strings in the output of the prosody module, which are already phonologically unique, are used to look up the phonetic transcriptions of the words in a dictionary. The accents are already present in these transcriptions. The prosodic boundaries are inserted, based on the annotation for the prosodic boundary. The resulting phonetic representation is passed on to "Spengi", which takes care of the eventual speech output.

In this application it is possible to use a dictionary of words, because the domain is restricted. If the domain would be more extended and would contain unpredictable words, using a dictionary will not be sufficient.

3.6.2 Phrase concatenation

The other form of speech synthesis is concatenating speech recordings of segments of the sentences. For both accented and unaccented words a different speech sample has been recorded. The presence of prosodic boundaries also is of influence on the pronunciation of words. For example, syllables preceding a pause are lengthened. Each possible appearance of a word or phrase has to be recorded. Each recorded fragment is labelled with a string as represented in the enriched text.

To form sentences, GoalGetter tries to match as big as possible pieces of the sentence with strings associated with a recording. The recordings are concatenated and made audible as a complete utterance. For example, in the sentence:

Van "Dijk deelde "vier "gele "kaarten uit.///

Van "Dijk is matched to the label of the file containing the accented version of the proper name *Van Dijk*.

Chapter 4: The speech material

4.1 Introduction

To investigate how speakers use accentuation, natural speech had to be collected. This was done by inviting people to read some texts aloud. This speech was recorded in a studio at the IPO. Afterwards, *expert listeners* were invited to listen to this material and indicate on which words they heard accents. This study is based on these data.

4.2 Creation of the speech data

4.2.1 The texts

The texts were created using GoalGetter. Some example teletext pages were created by hand. These were input to GoalGetter. In the current implementation of GoalGetter, the user is presented with several intermediate steps of the generation process. Two of them were recorded, the prosodically annotated text and the plain text derived thereof. The plain texts were used for obtaining the natural speech from speakers, the annotated versions were used for comparison.

In total, eleven texts were selected from the texts generated. One text was used as a training text for the speakers. All speakers had to read the same ten texts.

To get every speaker to read the numeric expressions in the same way they were kept in text format, just like GoalGetter produced them. Presenting those numeric expressions in digits introduces the possibility that speakers use different ways to pronounce them. For example: '2340' can be pronounced like 'twee duizend drie honderd veertig' (two thousand three hundred and forty) or 'drientwintig honderd veertig' (twenty three hundred forty).

4.2.2 Recording

For the recordings, seven people were invited to read the generated texts aloud. They were all working at the IPO on a temporary basis as graduate students. There were three female and four male speakers.

The human speech was recorded. The texts were presented to the speakers on paper, accompanied by an instruction. Appendix 4A contains the full instruction the speakers received. Each text had to be read silently before reading it aloud. This was required to prevent the speakers from reading the texts sentence by sentence, thus leaving out cross-sentence relations.

When a mispronunciation was made, or another mistake, for example, a hesitation, the speaker was asked to read the story containing the mistake again at the end of the session, because these mistakes influence the placement of accents.

The first text was a test text for the speaker to get acquainted with the task, and it was used to tune the recording equipment.

4.3 Extracting the accentuation data

4.3.1 Selecting the material

First, the material by three of the seven speakers was excluded. These contained too many mistakes or were read too fast, therefore they were not spoken fluently.

Secondly, from the texts of the remaining four speakers a selection had to be made, since presenting all the texts to listeners would cost them too much time. Text 1 and 2 were excluded first. From the fact that most speakers made one or two mistakes in these texts, it seemed that the speakers needed more than just the test text to get comfortable with the task. From the remaining eight texts, texts 3,4 and 5 were selected. These texts contained examples of all of the syntactic constructions present in the speech material.

The selected speech material was stored on a hard-disk, sampled down to 16 kHz. Furthermore, it was cut into separate sentences to make the task of the listeners more convenient.

4.3.2 Expert Listeners

Nine people from the Speech Group of the IPO were invited to listen to the speech data. They were considered to be *expert listeners.* This means that they have knowledge of prosodics. They were given scoring forms, containing the three texts, to indicate on which words they heard an accent. They were asked to make a distinction between a strong accent and a weak accent by assigning a "1" to the former and a "2" to the latter. They assigned nothing to words without an accent. Appendix B contains the instruction the listeners received, Appendix 4A is an example of one page of the scoring form.

4.4 Flaws in the generation process

The texts created for this investigation are not as well-formed as may be necesarry for a natural assignment of accents by speakers. It has to be kept in mind that differences between human accentuation and the accentuation assigned by GoalGetter may be caused by flaws introduced by the text generation component.

In this subsection I will give a short inventory of the deficiencies encountered. These were expected to be taken care of during this research, so it was possible to manipulate the texts to improve the quality of the generated texts.

Summary of encountered flaws in text-generation:

- "later": recall section 3.4.5, the example sentences. I will quote them here again for clarity.

"a. in de 41e minuut scoorde Van Gastel voor Feyenoord.
b. in de 60e minuut scoorde de aanvaller Vos voor Feyenoord.
b'. 19 minutes later scoorde de aanvaller Vos voor Feyenoord.
c. in de 63e minuut scoorde Kluivert voor Ajax.

c'. 22 minuten later scoorde Kluivert voor Ajax."

As was mentioned in that section, sentence **c'**. should refer to sentence a., not to sentence **b'**. because the time contained in sentence **b'**. is presented implicitly. Therefore, expressing the information of sentence c. by sentence d.

d. 3 minuten later scoorde Kluivert voor Ajax.

is wrong. However, the current implementation of GoalGetter can not handle this in the correct way.

- redundant generation of team names: The name of a team is mentioned too often. In consecutive sentences the following may occur.

"X van PSV opende de score."
"X scoorde tien minuten later opnieuw voor PSV."
or:
"X van PSV scoorde in de 10e minuut."
"de PSV-speler X scoorde in de 20e minuut opnieuw."
or:
"In de 20e minuut kreeg X van PSV een rode kaart, dus moest PSV met tien man verder spelen."

- unnatural use of referring expressions: too often names are used to refer to persons presented before, where pronouns should have been used instead.

- Unnatural text structure will lead to accentuation that would not be encountered in natural speech. It may also deceive speakers and cause them to misinterpret the texts. This will lead to incorrect accentuation, too. An example of unnatural text structure is the generation of the same sentence constituent structure in two or even more consecutive sentences. For example, the following sequence of sentences could have been produced.

"in de 10de minuut scoorde Kluivert voor Ajax."
"in de 25e minuut scoorde Overmars voor Ajax."
"in de 40e minuut scoorde de aanvaller Obiku voor Feyenoord."

This sounds less then natural when it is read aloud.

4.5 Conclusion

The method followed leads to a collection of data on accentuation in the form of written 1's and 2's. These data represent the judgments of nine expert listeners on the place of accents produced by four speakers.

Chapter 5: Results

This chapter describes how the raw data on the score forms of the judges was processed in order to evaluate the differences in accentuation by the speakers and the GoalGetter system.

5.1 Processing the data

5.1.1 Introduction

The data that result from the experiment described in Chapter 4, are only available on paper in the form of written '1's' and '2's'. As mentioned in Chapter 4, the listeners were asked to assign a '1' to words on which they heard a 'strong' accent and a '2' to words on which they heard a 'weak' accent. If no accent was heard nothing had to be assigned to that word. From these score-forms the judgements of the listeners have to be made accessible for statistical investigation. For this purpose the statistical software package SPSS was used. Appendix 5A contains the SPSS syntax used in this chapter.

5.1.2 From analogue to digital data

The scores of the listeners were fed into the computer. For a word which received no annotation by a judge a '0' was inserted. If a word was judged to have a weak or strong accent a '1' or '2' was inserted respectively. For comparison the accents awarded to the words by GoalGetter were included. As the rules in GoalGetter only make a distinction between accented and unaccented, the accents are presented as strong accents.

Note that, now, the labelling of the accents is reversed in comparison to the scores given by the listeners. This is done for practical reasons. It seems more logical to assign a higher digit to a stronger accent. More accent corresponds to a bigger number.

5.1.3 Organizing the data

Because one of the goals of the research is to find out on which words the accentuation by speakers and by the prosody module of GoalGetter is different, the most meaningful way to split up the data is by word. There are 211 words in the texts. Each word is spoken by four speakers. These utterances are each judged by nine listeners. This results in a number of 7596 judgements, corresponding to the same number of cases.

The raw data were read into SPSS, resulting in a data matrix as presented in figure 1. In the upper row, the variables are represented. 'wordnr' is the number of the word under consideration and ranges from 1 to 211; 'speaker' represents which speaker has uttered the word and ranges from 1 to 4; 'judge' indicates which of the nine listeners has judged the word; 'score' gives the judgement of this listener on the word uttered by the speaker, possible values are 0, 1 or 2; finally, 'pros' represents the accentuation predicted by the rules of GoalGetter, it only takes the values 0 or 2.

wordnr	speaker	judge	score	pros
1	1	1	0	0
2	1	1	1	2
....	....	....	....	....
211	4	9	2	0

Figure 1: a part of the SPSS data matrix.

5.1.4 Classifying the variables

Looking at the data, the dependent and independent variables can be easily distinguished. The independent ones are 'speaker' 'judge' and 'wordnr', the dependent variables are 'score' and 'pros'. 'speaker', 'judge' and 'wordnr' are nominal variables. 'score' and 'pros' are more difficult to classify. Maybe, 'score' can be said to be an ordinal variable, it ranks the cases according to "not accented", "weakly accented" or "strongly accented". In this light, 'pros' has to be considered to be an ordinal variable too, but in this case there are only two classes.

5.2 A perceptual measure for accentuation

The final objective is to find cases in which the speakers do not behave in the way the accentuation rules of GoalGetter predict. But first a perceptual measure for the accentuation by the speakers has to be found.

By putting the variables 'score' and 'judge' in a crosstable, using 'speaker' as the control variable, a global overview of the distribution of the scores can be obtained. The results are summarized in table 1. In this table, for each judge the frequency of observations is listed, subdivided by speaker. The final column gives the total frequencies per speaker and the last three rows give the total results per judge.

Table 1 shows that, with respect to the discrimination of weak and strong accents, the judges seem to have a different approach. For example, from the material of speaker 1, judge 1 and 2 both judge 119 words to be unaccented, but judge1 judges 45 words to be weakly accented while judge2 judges only 14 words in that way. Overall, judge4 only judges 17 words to be weakly accented against his colleagues at least more than double. Four of them even say that more than 100 words are accented weakly.

It is remarkable however that the number of times they hear no accent shows relatively little deviation. This might indicate that on the distinction between 'no accent' and 'accent' they do agree.

Appendix 5B contains bar charts of these figures. Similar observations can be made on these graphical representations.

Table 1: the frequency of observations for each judge, subdivided by speaker, together with the total percentages per speaker (last column) and the total frequencies per judge.

		judge									
speake r	**score**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**tot%**
1	0	119	119	120	119	120	105	107	116	105	54.2
	1	45	14	30	6	40	51	20	27	14	13.0
	2	47	78	61	86	51	55	84	68	92	32.8
2	0	112	109	107	119	118	105	103	107	105	51.9
	1	45	7	20	3	34	43	16	20	13	10.6
	2	54	95	84	89	59	63	92	84	93	37.5
3	0	111	106	107	115	115	100	90	103	104	50.1
	1	44	10	31	4	46	49	19	19	15	12.5
	2	56	95	73	92	50	62	102	89	92	37.4
4	0	105	98	103	106	106	95	94	98	98	47.6
	1	33	7	20	4	55	52	12	19	6	11.0
	2	73	106	88	101	50	64	105	94	107	41.5
total	0	447	432	437	459	459	405	394	424	412	50.9
	1	167	38	101	17	175	195	67	85	48	11.8
	2	230	374	306	368	210	244	383	335	384	37.3

Based on these findings, a new variable 'posscore' ("positive score") is computed, merging categories 1 and 2 of 'score' to one category "accented". The resulting data-matrix looks like figure 2.

wordnr	speaker	judge	score	pros	posscore
1	1	1	0	0	0
2	1	1	1	2	2
....	....	....	....	....	....
211	4	9	2	0	2

Figure 2: a part of the adapted data-matrix.

5.2.1 Agreement between listeners

To test if the listeners agreed on the distinction between "accent" and "no accent", kappa's coefficient of agreement, K, was used. This test was developed for investigating nominal variables. The variable to test, 'posscore', only makes a distinction between accented and not-accented. It is difficult to say that this is an ordering of some sort. Therefore, 'posscore' can best be treated as a nominal variable.

The kappa-coefficient is the ratio of the proportion of times that the judges agree (corrected for chance agreement) to the maximum proportion of times that the judges could agree (corrected for chance agreement). (Siegel & Castellan, 1988). K=0 means that there is no agreement, K=1 signifies total agreement.

For 'posscore' the value of K=0.827 and the calculated z=143.8 (for N=7596 words, k=9 listeners and m=2 categories). This is significant at the .01 level. So, because the kappa-coefficient indicates a high agreement we may conclude that the listeners have treated the task in a similar way.

It is important to see that cases have to be separated by 'speaker'. Of course a speaker has to obey rules to produce a natural sounding message, but he has a great deal of freedom producing an accent and there is always a chance that he produces an error. Therefore, it will not be useful to calculate agreement between speakers and it will be wrong to take the mean of the observed accentuation for a group of speakers.

5.2.2 Observed accentuation

Now that it is demonstrated, in section 5.2.2, that the listeners treat the task in the same way, we can define a new variable, 'totscore' ("total score") representing perceptual or observed accentuation of a word, as the mean of the scores of the nine judges.

A new data matrix is computed, eliminating the variable 'judge', creating the new variable 'totscore'. This new variable gives the mean score for each word uttered by each speaker. 'totscore' ranges from 0.00 to 2.00 in steps of 1/9, due to the fact that it is a mean value for the 'posscore' of nine listeners. Figure 3 is an example of the new data matrix. 'totpros' is the mean value for 'pros' and therefore the predicted accentuation of GoalGetter, which will always be equal to 'pros'. The data present in this matrix will be used in section 5.3.2 for computing differences between the observed and predicted accentuation.

wordnr	speaker	totscore	totpros
1	1	0.00	0.00
2	1	1.78	2.00
....	....	....	....
211	4	1.78	0.00

Figure 3: a part of the derived data matrix.

For each word uttered by each speaker, we now have a measure representing accentuation observed in natural speech and one representing the accentuation assigned by GGPC.

5.3 Computing the differences

On first inspection of the overall results of the observations and the prediction of the system there seems to be no big difference. For illustration, compare tables 2 and 3.

Table 2. The frequencies of the observations of "accent" (posscore=2) and "no accent" (posscore=0) in % per speaker.

	posscore	
speaker	0	2
1	54.2	45.8
2	51.9	48.1
3	50.1	49.9
4	47.6	52.5
total	50.9	49.1

Table 3: The frequencies of "accent" (pros=2) and "no accent" (pros=0) predicted by GoalGetter, in %.

pros	
0	2
55.5	44.5

From table 2, it seems that in about 50% of the cases words are judged to be unaccented both for individual speakers and overall, the number of predicted unaccented words comes close to these figures, it is just a little higher, 55,5%. This indicates that the rules of the system are probably a good prediction for the assignment of accents.

5.3.1 The computation

Now that a perceptive measure has been found, the difference between the observed and predicted accentuation can be computed. The result of this computation is represented by the new variable 'differ' (difference). A negative value for 'differ' means that there is no accent observed while GoalGetter predicts one, a positive value means that listeners do hear an accent while GoalGetter does not predict one. Now, it is fairly simple to filter out only those cases where a difference occurs. The result of the filtering is represented in a table in Appendix 5C.

From these results it has to be decided which differences are significant. In other words, it has to be decided how many judges have to agree. For this decision, a test with the binomial distribution is used. The zero hypothesis is that judgements are given at random. The population, n, consists of 9 judges, and the chance of failing the hypothesis, α, equals 0.05 (corresponding to a significance level of 95%). In this case a two-tailed test has to be performed, because both the accentuation and the absence of accentuation of words must be tested. This test shows that if one or no listeners hear an accent, it can be said with 95% certainty that no accent is produced on that particular word and if eight or more listeners hear an accent, with 95% certainty one will have been produced.

The consequences of these results are that if 'totscore' is equal to or less than 0.33 no accent will have been produced. If 'totscore' equals or is bigger than 1.78 an accent will be produced. The significant differences computed in the variable 'differ' will therefore be 2.00 and 1.78, signifying an observed accent where none was predicted or -1.78 and -2.00, an accent predicted but not observed.

5.3.2 The differences

At this point the factor 'speaker' plays a big role. Speaker 2 is considered best by most of the listeners. They were asked to rank the speakers according to naturalness of speech style. This includes speaking rate and hesitations.

For 'differ' <= -1,78 or 'differ' >= 1,78 for speaker 2 the following list of word numbers is found: 35, 38, 119, 169, 172, 175 and 181. These seven words occur in four different sentences. For all these words, as is clear from Appendix 5C, the other speakers also show some deviation from the predicted accentuation, although sometimes less rigorous. This is an indication that for these words the difference could be realistic.

Eventually it shows that only few differences can be found. From the 211 words 84 cannot receive an accent in Dutch, for example because they are prepositions, so 127 possible accent carrying words remain. Only 7 of them are indeed accented differently than predicted, this is 5.5%.

5.3.3 The context

The seven words mentioned in section 5.3.2, are found in the following sentences; the first digit is the number of the text, the second one is the number of the sentence. Words in bold are not accented by the GoalGetter system while they are by the speakers, for words in italics the opposite holds (/'s indicate prosodic boundaries; an " in front of a word indicates that that word is accented):

1.5 Van **Vliet** deelde "geen *"rode* "kaarten uit.///

3.2 "Vijf en veertig honderd "bezoekers / bezochten het **Gemeentelijk** "Sportpark.///

3.6 In de "tachtigste minuut / kreeg "Cocu een "rode "kaart // dus moest **PSV** met "tien *"man* "verder spelen.///

3.7 **Vlak** voor het "eindsignaal / bepaalde de %**PSV** speler "Jonk de "eindstand / op "@twee // - "@twee.///

In Appendix 5D the complete listing of all texts is given including the annotation of the GoalGetter system.

5.4 A second experiment

Two cases: "Gemeentelijk" and "Vlak" can be easily explained. On the other cases however, nothing really can be said in this stage, based on so little data. Therefore an additional experiment is performed concentrating on these constructions. "Gemeentelijk" and "Vlak" will be explained in chapter 6.

5.4.1 Design

The second experiment is directed towards the three constructions not yet explained above. Material already recorded for the first experiment can be used. Only speaker 1 and 2 are selected. They are judged by the listeners to have a more natural speaking style than the other two speakers. This also reduces the amount of utterances to be presented to the judges. Two texts are selected from the material. From one text the "introduction" and "body", see chapter 3 on GoalGetter, are used, from the other only the "conclusion" is used.

What remains is a composed text containing 76 words, spoken by two speakers. The texts are cut into sentences and presented to the same nine listeners as in the first experiment. The instructions are the same as in the first experiment.

5.4.2 The data and results

The data were processed as described in sections 5.1 and 5.2. The results are given in table 4.

Table 4:

wordnr	speaker 1	speaker 2	woord
39	2.00	2.00	Sparta
42	-1.78	-1.78	man
65	2.00	2.00	Dijk
68	-1.78	-1.56	rode

The numbers of the words on which differences in accentuation appear are 39, 42 and 65. Also word 68, "rode" is kept in mind, because it is present in the first experiment in the same construction.

4.1 "Sparta nam in de "achtste "minuut de "leiding / door een "treffer van de % "verdediger de "Bruin///

4.2 in de "zevenenveertigste minuut / scoorde van der "Laan van Sparta///

4.3 "negenentwintig minuten "later / kreeg "Hansma een "rode "kaart // dus moest **Sparta** met "tien "*man* "verder spelen///

4.4 in de "tachtigste minuut / bepaalde de % "aanvaller "Fortes de "eindstand / op "@vier // -

"@twee///

4.5 "scheidsrechter van "Dijk / "leidde het duel///

4.6 van **Dijk** deelde "geen "*rode* "kaarten uit///

4.7 hij deelde "vijf "gele kaarten uit///

5.4.3 Comparison with first experiment

Comparing these differences to the ones from the first experiment we see the following resemblances. In sentence 3.6 and 4.3 both "PSV" and "Sparta", in the same position in the sentence, should be accented according to the speakers. Also in sentence 4.3, "man" should be de-accented just as in sentence 3.6. "rode" in sentence 4.6 is de-accented, same as in the first experiment. "Dijk" in sentence 4.6 should be accented, just like "Vliet" in sentence 1.5.

Chapter 6: Ameliorations for GoalGetter

In chapter 5, several cases were found where the observed accentuation differed from the predicted accentuation. In this chapter I will try to explain those differences considering their context and syntactic structure and if possible, I will try to make a proposal for changes to the rules of the system.

6.1. Introduction

The differences found in the two experiments described in chapter 5 are summarized in the following listing. Words in bold were observed to be accented, while no accent was assigned by the system. The ones in italics are observed to be unaccented, while an accent was assigned by the system.

1.5 Van **Vliet** deelde "geen *"rode* "kaarten uit.///
3.2 "Vijf en veertig honderd "bezoekers / bezochten het **Gemeentelijk** "Sportpark.///
3.6 In de "tachtigste minuut / kreeg "Cocu een "rode "kaart // dus moest **PSV** met "tien *"man* "verder spelen.///
3.7 **Vlak** voor het "eindsignaal / bepaalde de %**PSV** speler "Jonk de "eindstand / op "@twee // - "@twee.///
4.3 "negenentwintig minuten "later / kreeg "Hansma een "rode "kaart // dus moest **Sparta** met "tien "*man* "verder spelen///
4.6 van **Dijk** deelde "geen "*rode* "kaarten uit///

The complete texts containing these sentences are listed in Appendix 5D.
In the following part of this chapter these cases will be discussed in detail. Each case will be discussed in a separate section. First, I will analyse the case and secondly I will give a proposal for solving it.

6.2. Case 1: "Vlak" and "Gemeentelijk"

3.7 **Vlak** voor het "eindsignaal / bepaalde de %PSV speler "Jonk de "eindstand / op "@twee // - "@twee.///
3.2 "Vijf en veertig honderd "bezoekers / bezochten het **Gemeentelijk** "Sportpark.///

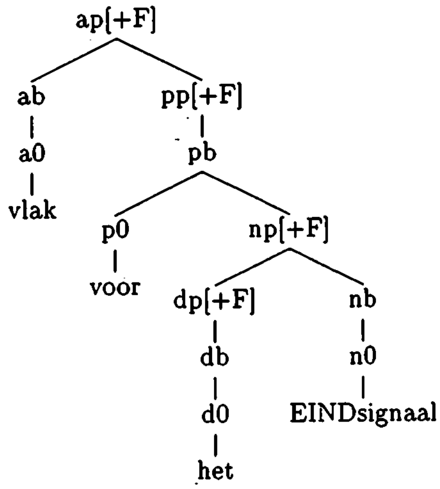
6.2.1 Analysis

These two cases can be explained as deficiencies of the constructed trees. A simple analysis of "Vlak" shows that the sub-tree in which it appears is incorrectly constructed. In this representation the phrase is declared as an adjective phrase with "vlak" as the head. In fact, the phrase "vlak voor het eindsignaal" is a prepositional phrase and "voor" is the head. One test we have performed by leaving "vlak" out and investigating if the rest of the phrase should be left out, too. This is not the case, "voor het eindsignaal" can still be used without creating a syntactically wrong sentence. If "vlak" would have been the head of the phrase this would not be possible. "vlak" is the head (and only constituent) of an adjectival phrase which is part of the prepositional phrase.

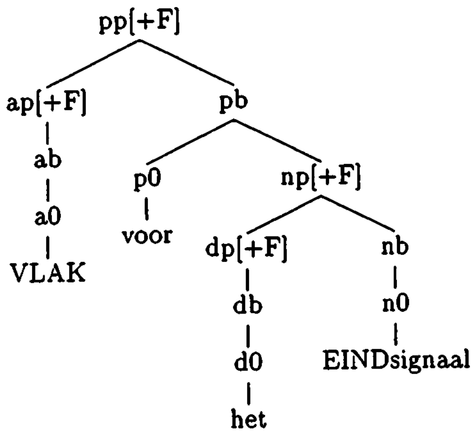
"Gemeentelijk" occurs in the database as a list of two strings: {"Gemeentelijk", "Sportpark"}. This has a consequence that the name is treated as a proper name consisting of two strings. The accent is placed on the second part, because a default structure is assigned to it. This is sufficient for names like "de Kuip" or "de Meer", but in this case the structure of the name is richer and has to be accounted for too. The whole expression should be represented as a phrase with internal structure.

6.2.2. Proposed changes

These two faults can be easily solved. The sub-tree given in the relevant templates looks like:



In this tree structure, the ap-node launches an accent. This accent is distributed along the strong sons and eventually lands on "eindsignaal". This structure should be changed according to the syntactic function of "vlak" into:

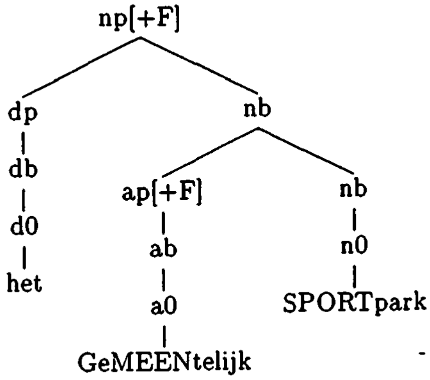


Now the accent launched at the ap-node is distributed to "vlak".

The absence of an accent on "Gemeentelijk" can easily be solved by defining an internal structure for the compound name and adding this to the lexicon. The internal structure should be defined as the normal structure for a noun phrase consisting of an article, an attributive adjective and a noun;

```
np-[dp-[db-[d0-"het"]],
   nb-[ap-[ab-[a0-"Gemeentelijk"]],
       nb-[n0-"Sportpark"]
      ]
   ]
```

represented as a tree:



In this structure "Gemeentelijk" gets an accent because the dominating ap-node launches one.

6.3. Case 2: "PSV" and "Sparta"

3.6 In de "tachtigste minuut / kreeg "Cocu een "rode "kaart // dus moest **PSV** met "tien *"man* "verder spelen.///

3.7 **Vlak** voor het "eindsignaal / bepaalde de %**PSV** speler "Jonk de "eindstand / op "@twee // - "@twee.///

4.3 "negenentwintig minuten "later / kreeg "Hansma een "rode "kaart // dus moest **Sparta** met "tien *"man* "verder spelen///

6.3.1 Analysis

In the analysis of these cases the context plays an important role. The previous sentences are respectively:

3.5 "twee minuten "later / scoorde de % "PSV speler "Vink///
4.2 "in de "zevenenveertigste minuut / scoorde van der "Laan van Sparta///

In these sentences both "PSV" and "Sparta" are used in a different manner than in 3.6 or 4.2. In 3.5 "PSV" is part of the compound "PSV-speler". The whole phrase, identifying the player, "de PSV speler Vink", is the subject of the sentence, while in 3.6 "PSV" itself is subject of the subordinate clause ".. en dus moest PSV met tien man verder spelen". In 4.2 "Sparta" is part of a prepositional phrase which says something about the player named "van der Laan". The complete phrase "van der Laan van Sparta" is subject of 4.2. In 4.3 "Sparta" is subject of the subordinate clause. Both in sentences 3.6 and 4.3, "PSV" and "Sparta" are head of a phrase, while in the previous sentences they were just modifying expressions. These syntactic differences may prevent the second appearances of the team name to be deaccented by the speakers. GoalGetter assigns no accents these words because they are considered "given information", cf. section 2.4. GoalGetter takes the objects in the second sentences as being equal to the objects in the previous sentences. Perhaps that is not the correct way of treating these words. The different syntactic status of the objects across sentences may cause speakers to not deaccent the second appearances. Corresponding to the different syntactic status, there is also a semantic difference. The object of the team name contributes to the concept of the individual player. The concept of the team name is therefore subordinate to this concept. In the next sentence the concept of the team is independent and appears at the same semantic level as the player concept.

For "PSV" in sentence 3.7 a similar explanation holds. In this sentence "PSV" is used in the same construction as in 3.5 and "PSV" in the previous sentence is subject of the subordinate clause.

Only in these three cases these phenomena occur. It is very difficult to draw conclusions based on so little material. However, there are no examples in these texts that contradict the analyses. Cases where the team name is used in a syntactically similar way in consecutive sentences, are not present in the data, so nothing can be said about them.

Another explanation may be found in a contrastive function of the accentuation of the team name. By placing an accent on the team name it is contrasted from the unmentioned , but known, opponent. Although the opponent is unmentioned in a sentence, it is in some way always present, because there are only two teams in one match. Based on the data available, it can not be decided which is the correct explanation.

6.3.2 Proposed changes

Considering the syntactic analysis of section 6.3.1, the following changes can be proposed. Before a word is deaccented, check if two similar words in two consecutive sentences are used at the same conceptual level. If concept 1 is part of a bigger concept which plays a role at the predicate/argument level and concept 2 plays a role at this level itself, the second appearance of the word should not be deaccented. How to implement this is again another question.

The second analysis leads to a rule that marks the names of the teams. In that case, the deaccentuation rules should make an exception for team names.

6.4 Case 3: "van Vliet" and "van Dijk"

1.5 Van **Vliet** deelde "geen *"rode* "kaarten uit.///
4.6 van **Dijk** deelde "geen *"rode* "kaarten uit///

6.4.1 Analysis

The previous sentences are respectively:
1.4 "scheidsrechter van "Vliet / "leidde het duel///
4.6 "scheidsrechter van "Dijk / "leidde het duel///

So, both names are given information and no accents are assigned to them. Both expressions "scheidsrechter van Vliet" and "van Vliet" refer to the same concept and even have the same grammatical function. A possible explanation is that usually in real, fluent conversations speakers do not use the same name in two consecutive sentences, at least not in the context described here. They rather refer to the person in the previous sentence by a personal pronoun or another referring expression. Only in cases where a special focus or contrast is needed or wanted will the same name, which will be accented, be used. So, in these cases, the difference in accentuation is caused by the system's wrong choice of referring expression.

These differences, combined with the findings in section 6.3, may also be an indication that all names, proper names and names of the teams, always have to be accented.

6.4.2 Proposed changes

A possible solution will be to implement a restriction that prohibits the use of proper names in two consecutive sentences and that instead forces the use of alternative referring expressions, such as personal pronouns in these cases.

If this rule is implemented optionally, another rule must be implemented that prevents the deaccentuation of proper names. However, to me, it seems better not to change or add rules that have no theoretical founding, based on so little data from such a limited domain.

6.5 Case 4: "rode" and "man"

1.5 Van Vliet deelde "geen *"rode* "kaarten uit.///
3.6 In de "tachtigste minuut / kreeg "Cocu een "rode "kaart // dus moest PSV met "tien ***"man"*** verder spelen.///
4.3 "negenentwintig minuten "later / kreeg "Hansma een "rode "kaart // dus moest Sparta met "tien ***"man*** "verder spelen///
4.6 van Dijk deelde "geen ***"rode*** "kaarten uit///

6.5.1 Analysis

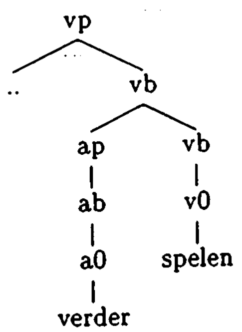
At first, these two cases seem similar, of three consecutive words that are assigned an accent by GoalGetter, the middle word is left unaccented by the speaker. In this section I will analyse these cases separately

6.5.1.1 Case 4a: "met tien man verder spelen"

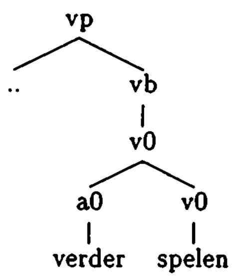
The first thing to inspect is the verb phrase "verder spelen" in order to construct the correct tree structure. In this phrase "verder" is a part of the verb rather than an adjunct.
Looking at the following sentences we can see that "verderspelen" behaves like "naspelen" and differs from "goed spelen". So "verderspelen" is, like "naspelen" a verb with a separable prefix.

(6.1): omdat hij goed wilde spelen
*omdat hij wilde goed spelen
(6.2): omdat hij het na wilde spelen
omdat hij het wilde na spelen
(6.3): omdat hij verder wilde spelen
omdat hij wilde verder spelen

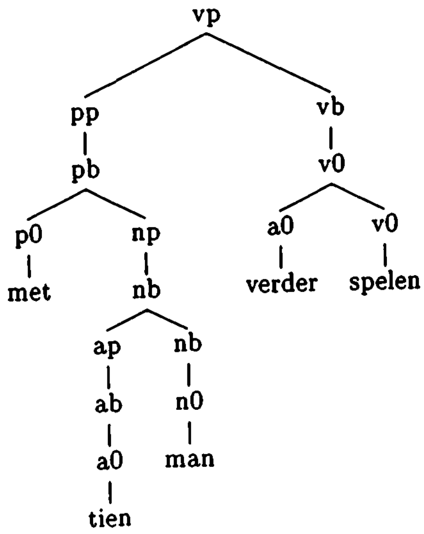
So the sub-tree of "verderspelen" should not be constructed as:



but as:



The tree-structure for the complete verb phrase is the following:

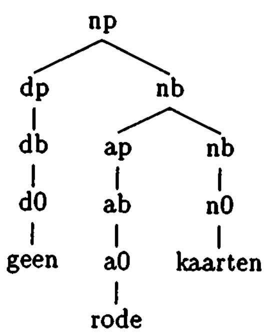


Does the Rhythm Rule apply here? For the definition of the Rhythm Rule see section 2.5.

In this case, the A node is "tien" (ap), the B node is "man" (nb) and the C node is "verder spelen" (the upper v0). If the Rhythm Rule has to apply here, the compound as a whole should be seen as the C node. The nb-node ("man") now becomes weak, so no accent will be distributed along the branch between np and nb. The ap-node becomes strong. This will have no visible effect because this node launches an accent of its own. So the Rhythm Rule would account for an accentuation according to the observed accentuation.

6.5.1.2 Case b: "geen rode kaarten"

In this example the difference in accentuation is caused by the fact that the ap-node launches its own accent. Besides that, the determiner, which receives an accent launched by the dominating dp-node, is able to carry an accent too. This results in three accents close to each other, causing a clash of accents. Comparing the tree structure in the definition of the Rhythm Rule with the tree structure of this phrase shows that the rule does not apply here.



At the word-level the sequence of accents assigned by the system is [accent-accent-accent]. According to what speakers produce, the word in the middle should be deaccented, [accent-no accent-accent]. It seems that the only way to solve this problem is assuming a low-level linear rhythmical effect.

Baart, 1987, shows that a low-level linear explanation of rhythmical phenomena that only considers the words without looking at the syntactic structure, does not suffice. It should be noted however that in his refutation of the linear approach, Baart, 1987, only considers cases with a weak-strong-strong configuration, resulting in [no accent-accent-accent].

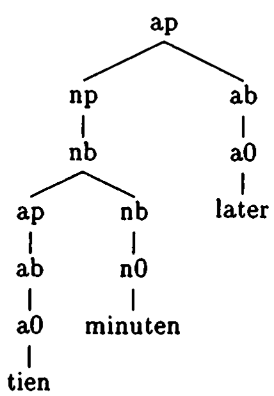
In this case the accents all occur within one phrase. What happens when the accented words are not part of one phrase can not be investigated because no such cases are present in the data, ([accent-accent]-[accent] -> [accent-no-accent]-[accent]???). Furthermore, situations in which there are four accents in a row (either in one phrase or divided over two phrases) do not occur in the data, ([accent-accent-accent-accent] or [accent-accent]-[accent-accent]).

6.5.2 Proposed changes

6.5.2.1 Case 4a: "met tien man verder spelen"

A solution would be to implement the Rhythm Rule in the prosodic component of GoalGetter. This rule seems to apply here, taking into account that the verb is a compound and that the node dominating the constituents is to be considered the relevant lowest node of the verb.

Implementation of this rule could affect other constructions with similar syntactic structure. There is one such phrase, "tien minuten later". This phrase frequently occurs in expressing the time of certain events, of course with different numeric values. The syntactic sub-tree is constructed as:



If the Rhythm Rule would be implemented it could be applied to this tree. However, in the sentences considered, "minuten" is always deaccented because it is given information. Every sentence in which "<nr> minuten later" occurs is preceded by a sentence in which the word "minuut" or "minuten" is mentioned. Based on the available data, it can only be concluded that when "minuten" is observed to be deaccented, GoalGetter assigns no accent to this word, but nothing can be said about the reason; deaccentuation of 'given information' or application of the Rhythm Rule.

It is possible that a sentence containing the phrase "<nr> minuten later" follows a sentence without "minuut", for example the sentence: "vlak na rust ...." (shortly after the break .....). Only in this case, implementation of the Rhythm Rule will directly affect the distribution of accents. This sequence of sentences is not present in the data, however.

6.5.2.2 Case b: "geen rode kaarten"

A solution could be to implement a rule that takes care of cases in which three words in a row receive an accent. The middle accent should be deleted. This can be done on different levels. One is to formulate a rule that detects the presence of three adjacent strong leaves and converts the middle node to weak (*s s s* -> *s w s*). Now this node can not receive an accent.

Another way to realize this is to formulate the rule on word level. If three adjacent words receive an accent, the accent on the middle word is discarded.

There are not enough data to investigate all possible forms of accent clash. Therefore it is not possible to formulate a rule that covers all cases and it will be better to choose the least intervening solution. This is the second one, which does not affect the system of accent distribution, it only works at the articulatory level. Problems with pronouncing three accents so close to each other are seen as the reason for deaccentuation.

In both cases the rule should only apply when the words contain few syllables. Perhaps it is better to define the rule in terms of syllables. Again, there are not enough data to formulate conditions on the number of syllables.

6.6 Conclusion

In this chapter I have analysed the differences in accentuation by the speakers and by GoalGetter. Furthermore, I have attempted to give a solution to avoid these problems.

Two of the cases, that of "Gemeentelijk" and of "Vlak", were minor problems due to incorrect definitions of the syntactic structure of the constituents in which they occurred. Proposals for solving them have been given. Two other cases, that of PSV/Sparta and Van Dijk/ Van Vliet, seem to be in the field of deaccentuation. Suggestions for solutions have been given for these, too.

Chapter 7: Conclusions

In GoalGetter, accents are assigned by GoalGetter's Prosody Component (GGPC), which is based on Focus Accent Theory and augmented with an implementation of a discourse model as proposed by Van Deemter (1993).

The goal of the research described in this paper was to evaluate the rules for the assignment of accents implemented in the GoalGetter application. For this purpose, I compared the assignment of accents by human speakers to the assignment of accents by GoalGetter. A second objective was to find out whether it is possible to distinguish different levels of strength of accents. Therefore, the listeners were asked to assign a '1' to a strong accent and a '2' to a weak accent.

The listeners were not able to make a distinction between weak and strong accents in a consistent way. However, because listeners seemed to judge unaccented words consistently, I was led to believe that listeners could make a distinction between 'accented' and 'not accented'.

Therefore, judgements of 'weak' and 'strong' accents were merged into one category, 'accented'. After restructuring the data in this way, only few cases were found on which a difference between the assignment of accents by GoalGetter and the accentuation as judged by the listeners occurred.

Based on this, two conclusions can be drawn, (1) that it is not useful to distinguish between 'weak' and 'strong' accents, because listeners use different criteria for judging them; (2) that the rules for accentuation implemented in GGPC are a good approximation of the way the speakers use accents.

The results of this investigation are valid only for this application, because the domain is limited and because of the fact that accentuation by speakers may have been influenced by the structure of the texts. The sentences of the texts that had to be read, were generated by GoalGetter's Text Generation component which introduced a lot of unnatural formed texts, for example, because the same construction was used in a series of consecutive sentences. This was solved as much as possible by selecting texts by hand.

In addition, the results are limited inside this domain, because it is possible that not all syntactic constructions are present in the speech material, which is caused by the limited amount of samples. Only relatively little material could be judged by the listeners due to time limitations. Therefore, the selection of texts had to be performed carefully to contain as much variety in syntactic constructions as possible, because it was not possible to predict in advance on which constructions or words differences would occur.

Some general conclusions can be drawn. The first conclusion is that a correct generation of texts, or knowledge of the structure and coherence of texts is necessary for a correct assignment of accents. The second one is that the amount of material used in this investigation is only sufficient to get a notion of the occurring phenomena and that for research in this area a larger pool of data is needed.

For the differences that did occur, I tried to find an explanation and proposed solutions. This was difficult due to the limited amount of data. Two cases were caused by incorrect definition of the syntactic structure of a constituent (on the words "vlak" and "Gemeentelijk"). Two other cases can most likely be ascribed to flaws in the text generation component ("Van Dijk"/"Van Vliet" and "PSV"/"Sparta"). The last two cases can be related to rhythmic factors (in the phrases: "met tien man verder spelen" and "geen rode kaarten"). One of them is due to the absence of the Rhythm Rule, the other seems to point in the direction of some kind of low level rule that prevents pitch accent clash.

References

Baart J. L. G. (1987), "Focus, Syntax, and Accent Placement: towards a rule system for the derivation of pitch accent patterns in Dutch as spoken by humans and machines", Dissertation University of Leiden.

Chomsky N. & Halle M. (1968), "The Sound Pattern of English", Harper & Row, New York.

Deemter K. van (1993), "What's new? A semantic Perspective on Sentence Accent", in *Journal of Semantics, 11*, pp 1-31.

Dirksen A. (1985), "Accenting and deaccenting: A declarative approach", in *Proceedings of COLING-92*, pp. 865-869.

Gussenhoven C. (1983), "Focus, mode and the nucleus", in *Journal of Linguistics* 19, pp 377-417.

Halliday M. A. K. (1967), "Notes on transitivity and theme in English, part II, in *Journal of Linguistics* 3, pp 199-244.

Kiparsky P. (1966), "Ueber den deutschen Akzent", in *Studia Grammatica* VII, p 69-98.

Ladd R. D. jr. (1980), "The Structure of Intonational Meaning; Evidence from English", Indiana University Press, Bloomington.

--- (1983), "Even, focus and normal stress", in *Journal of Semantics* 2/2.

Lakoff G. (1972), "The global nature of the nuclear stress rule, in *Language* 48, pp 285-303.

Nespor M. & Vogel I. (1986), "Prosodic Phonology", Foris Publications, Dordrecht.

Schmerling S. F. (1976), "Aspects of English Sentence Stress", Texas University Press, Austin.

Selkirk E. (1986), "On derived domains in sentence phonology", in *Phonology Yearbook*, vol. 3, pp 371-405.

Siegel S. & Castellan Jr. N. J. (1988), "Nonparametric Statistics", second edition, McGraw-Hill Book Company, Singapore.

Terken J. & Collier R. (1995), "The Generation of Prosodic Structure and Intonation in Speech Synthesis", in *Speech Coding and Synthesis* (Kleijn W.B. and Paliwal K.K. eds.), pp 201-232, Elsevier Science Publischers B.V.

Terken J. & Nooteboom S. G., (1987), "Opposite Effects of Accentuation and Deaccentuation on Verification latencies for given and new information", in *Language and Cognitive Processes, 2*, pp 145-163.

Appendix

Table of contents:

Appendix to chapter 4

A: The instruction to the speakers

Instructie:
Lees iedere tekst eerst een keer door. Lees daarna het nummer van de tekst, wacht een paar seconden en lees daarna de tekst hardop voor. Stel je voor dat je een nieuwsbericht moet voorlezen. Mocht je je verspreken, lees dan eerst het stukje helemaal af, ga vervolgens door met de rest en aan het eind van de sessie kun je die tekst nog een keer voorlezen.

De eerste tekst is slechts bedoeld als een test om het opname-niveau in te stellen.

B. The instruction to the 'expert listeners'

Instructie

In de directory "/home/scratch/nachtegaal/voorlees.dat/" staan spraakfiles, die ieder een zin bevatten. Deze zinnen hebben betrekking op uitslagen van voetbalwedstrijden. Ze zijn door verschillende mensen uitgesproken.

Iedere uiting correspondeert met een zin op de volgende pagina's; de bestandsnaam van de uiting staat voor iedere zin.

Luister met de applicatie "soundfiler" naar iedere uiting en geef op papier in de bijhorende zin aan waar U accenten hoort. Noteer een "1", voor een sterk accent, of een "2", voor een minder sterk accent, voor (of door) het desbetreffende woord. Als U op een woord geen accent hoort hoeft U dus niks te noteren.
In sommige gevallen, op de lange getalsuitdrukkingen die in de deze teksten voorkomen, kan het zijn dat er meer dan een accent waargenomen wordt. Graag dit ook noteren.

Het totaal aantal uitingen is 88. Naar schatting neemt deze proef ongeveer 50 minuten in beslag. Neemt U aub de tijd en doe alles desnoods in 2 of meer sessies.

Dank voor Uw medewerking.
D. Nachtegaal
k 2.27

Translation:
"Instruction

In the directory "/home/scratch/nachtegaal/voorlees.dat/" speechdatafiles are saved, each containing a sentence. These sentences concern the results of soccermatches. They are recorded from different people.

Each utterance corresponds to a sentence on the next pages; the filename of the utterance is given in front of each sentence.

Listen to each utterance using the applicaton "soundfiler" and indicate in the corresponding sentence on paper where You hear accents. Give a "1" to a strong accent, or a "2", to a less strong accent, before (or across) that word. If You don't hear an accent on a particular word, don't score it.
In some cases, on the long numeral expressions present in these texts, it's possible that more than one accent can be heard. Please also indicate this.

The total amount of utterances is 88. This test will take about 50 minutes. Please take your time, if necesarry, do it in two or more sessions.

Thank you for your cooperation."
D. Nachtegaal
k 2.27

C: First page of the scoring forms.

The first page of the forms on which the listeners could indicate where they heard accents. "PB3_1.aiff" is the name of the file containing the speech data of the following sentence. It indicates that this file contains the data of the first sentence ('_1') of text 3 ('3') spoken by speaker B ('PB'). "PB5_9b" does not mean that it is a part of sentence 9. It contains the speech of sentence 10. This file was given this name because the internal ordening mechanism of the Silicon Graphics machine would place this file on the second position if its name would have been "PB5_10". In order to present the files in the right order I performed this simple trick.

(/home/scratch/nachtegaal/voorlees.dat/*)

Tekst 3:

PB3_1.aiff: Het duel tussen Vitesse en Ajax eindigde in nul - een.

PB3_2.aiff: Negenentachtighonderdzesentwintig toeschouwers bezochten Monnikenhuizen.

PB3_3.aiff: Litmanen scoorde in de drieentachtigste minuut voor Ajax.

PB3_4.aiff: Scheidsrechter van Vliet leidde het duel.

PB3_5.aiff: Van Vliet deelde geen rode kaarten uit.

PB3_6.aiff: Reuser en Schulp van Ajax kregen een gele kaart.

Tekst 4:

PB4_1.aiff: Feyenoord tegen Sparta eindigde in een - een.

PB4_2.aiff: Vijfentwintigduizend toeschouwers bezochten de Kuip.

PB4_3.aiff: Sparta nam na achtendertig minuten de leiding door een doelpunt van Van der Laan.

PB4_4.aiff: In de zeventigste minuut bepaalde Koeman de eindstand op een - een.

PB4_5.aiff: Scheidsrechter Luinge leidde de wedstrijd.

PB4_6.aiff: Van Bronckhorst van Feyenoord kreeg een gele kaart.

Tekst 5:

PB5_1.aiff: Het duel tussen Roda JC en PSV eindigde in twee - twee.

PB5_2.aiff: Vijfenveertighonderd bezoekers bezochten het Gemeentelijk Sportpark.

PB5_3.aiff: Roda JC nam in de drieenzestigste minuut de leiding door een doelpunt van de aanvaller

Babangida.

PB5_4.aiff: In de negenenzestigste minuut tikte Valckx de bal in het verkeerde doel.

PB5_5.aiff: Twee minuten later scoorde de PSV-speler Vink.

PB5_6.aiff: In de tachtigste minuut kreeg Cocu een rode kaart dus moest PSV met tien man verder spelen.

PB5_7.aiff: Vlak voor het eindsignaal bepaalde de PSV-speler Jonk de eindstand op twee - twee.

PB5_8.aiff: Scheidsrechter Van Hulten leidde het duel.

PB5_9.aiff: Klomp van PSV kreeg een gele kaart.

PB5_9b.aiff: Van de Luer van Roda JC kreeg ook een gele kaart.

Appendix to chapter 5.

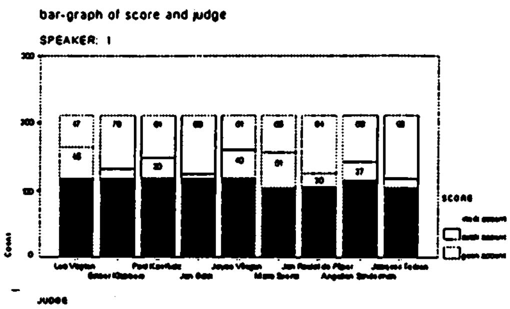
A: The listing of the SPSS-syntax file

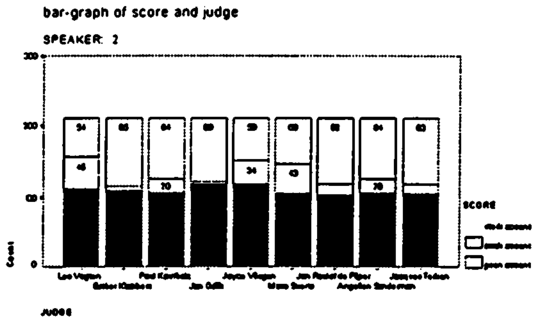
```
DATA LIST
  FILE='H:\spss\spsfile2.dat' FIXED / speaker 1-1  judge 2-2 wordnr 3-5 score 6-6 pros 7-7 .
EXECUTE.

/* creeer de kolom 'posscore', deze wordt bij sterk en zwak accent op 2
/* gezet.
COMPUTE posscore = 0 .
EXECUTE .

FORMATS posscore (f1.0) .
IF (score > 0) posscore = 2 .
EXECUTE .

/* enkele cosmetische toevoegingen om sneller verband te kunnen leggen
/* tussen de teksten en de opsomming van getallen.

/* creeer en vul de kolom text.
COMPUTE text = 0 .
EXECUTE .

FORMATS text (f2.0) .
IF ((wordnr > 0)  & (wordnr < 50)) text = 1 .
IF ((wordnr >= 50)  & (wordnr < 103)) text = 2 .
IF ((wordnr >= 103)  & (wordnr <= 211)) text = 3 .
EXECUTE .

/* creeer en vul de kolom 'sent'.
COMPUTE sent = 0 .
EXECUTE .

FORMATS sent (f2.0) .
IF ((wordnr > 0)  & (wordnr < 11)) sent = 1 .
IF ((wordnr >= 11)  & (wordnr < 19)) sent = 2 .
IF ((wordnr >= 19)  & (wordnr < 28)) sent = 3 .
IF ((wordnr >= 28)  & (wordnr < 34)) sent = 4 .
IF ((wordnr >= 34)  & (wordnr < 41)) sent = 5 .
IF ((wordnr >= 41)  & (wordnr < 50)) sent = 6 .

IF ((wordnr >= 50)  & (wordnr < 57)) sent = 1 .
IF ((wordnr >= 57)  & (wordnr < 64)) sent = 2 .
IF ((wordnr >= 64)  & (wordnr < 79)) sent = 3 .
IF ((wordnr >= 79)  & (wordnr < 90)) sent = 4 .
IF ((wordnr >= 90)  & (wordnr < 95)) sent = 5 .
IF ((wordnr >= 95)  & (wordnr < 103)) sent = 6 .

IF ((wordnr >= 103)  & (wordnr < 113)) sent = 1 .
IF ((wordnr >= 113)  & (wordnr < 121)) sent = 2 .
IF ((wordnr >= 121)  & (wordnr < 137)) sent = 3 .
IF ((wordnr >= 137)  & (wordnr < 150)) sent = 4 .
IF ((wordnr >= 150)  & (wordnr < 158)) sent = 5 .
IF ((wordnr >= 158)  & (wordnr < 175)) sent = 6 .
IF ((wordnr >= 175)  & (wordnr < 189)) sent = 7 .
IF ((wordnr >= 189)  & (wordnr < 195)) sent = 8 .
IF ((wordnr >= 195)  & (wordnr < 202)) sent = 9 .
IF ((wordnr >= 202)  & (wordnr < 212)) sent = 10 .
EXECUTE .

VALUE LABEL judge
  1 'Leo Vogten'
  2 'Esther Klabbers'
  3 'Paul Kaufholz'
  4 'Jan Odijk'
  5 'Joyce Vliegen'
  6 'Marc Swerts'
  7 'Jan Roelof de Pijper'
  8 'Angelien Sanderman'
  9 'Jacques Terken'.

/* bewaar dit in de spss-file 'cijfers.sav'.

SAVE OUTFILE='H:\SPSS\CIJFERS.SAV'.

/* toegepast op file: 'cijfers.sav'
GET FILE='H:\SPSS\CIJFERS.SAV'.

/*
/* KRUISTABELLEN
```

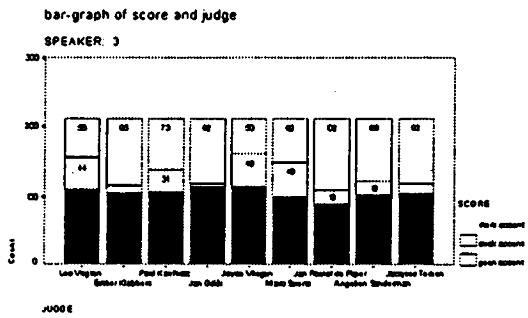
```
/*

/* kruistabel van luisteraar en score
CROSSTABS
  /TABLES=judge  BY score
  /FORMAT= AVALUE NOINDEX BOX LABELS TABLES
  /STATISTIC=ALL
  /CELLS= COUNT EXPECTED SRESID .

/* kruistabel van spreker en score
CROSSTABS
  /TABLES=speaker  BY score
  /FORMAT= AVALUE NOINDEX BOX LABELS TABLES
  /STATISTIC=ALL
  /CELLS= COUNT EXPECTED SRESID .

/* kruistabel van luisteraar en score per spreker
CROSSTABS
  /TABLES=judge  BY score BY speaker
  /FORMAT= AVALUE NOINDEX BOX LABELS TABLES
  /STATISTIC=ALL
  /CELLS= COUNT EXPECTED SRESID .

/*
/* STAAFDIAGRAMMEN
/*

/* staafdiagram van judge en score over alle sprekers
GRAPH
  /BAR(STACK)=COUNT BY judge BY score
  /MISSING=REPORT
  /TITLE= 'bar-graph of score and judge' 'taken from all speakers'.

/* staafdiagram van judge en score voor iedere spreker
USE ALL.
SORT CASES BY speaker .
SPLIT FILE BY speaker .
GRAPH
  /BAR(STACK)=COUNT BY judge BY score
  /MISSING=REPORT
  /TITLE= 'bar-graph of score and judge'.
  /FOOTNOTE= 'for each individual speaker'.
SPLIT FILE OFF .

/* deze functie creeert een nieuwe datafile, met daarin de oude kolommen;
/* 'speaker', 'text', 'sent' en 'wordnr'
/* nieuwe kolommen: 'totscore' en 'totpros'
/* deze bevatten het gemiddelde over alle luisteraars van respectievelijk
/* 'posscore' en 'pros' uit de datafile 'cijfer2.sav'
/* de nieuwe datafile creeeren met daarin de gemiddelde scores

AGGREGATE OUTFILE='H:\SPSS\SOMAGGR.SAV'
  /BREAK=speaker text sent wordnr
  /n_judges=N
  /totscore = MEAN(posscore) /totpros = MEAN(pros).

/*
/* dit stuk gebruikt de file 'somaggr.sav', de hierboven geaggregeerde
/* file.

GET FILE='H:\SPSS\SOMAGGR.SAV'.

/* het voegt er een kolom: 'differ' aan toe.
/* als differ positief is, dan betekent dat dat luisteraars wel een
/* accent horen en PROS-3 er geen maakt;
/* als differ negatief is, dan horen luisteraars geen accent en maakt
/* PROS-3 er wel een.

FORMATS aantal0 (F4) .
FORMATS aantal1 (F4) .
FORMATS aantal2 (F4) .

COMPUTE differ = totscore - totpros .
EXECUTE .

/* onderstaande syntax neemt uit een inputfile met daarin de kolommen:
/*   'speaker', 'text', 'sent', 'wordnr', 'totpros', 'totscore', 'n_cases',
/*  'differ'
/* de gevallen waarvoor geldt dat de accentuering significant verschilt van
/* de PROS-3 accentuering.
/* de inputfile is gegenereerd van de standaarddata file 'cijfers.sav'.
/* 'totpros' is de som van alle prosscores aangaande 1 woord van 1 spreker;
/* 'totscore is de som van alle posscores van alle luisteraars van 1 woord
```

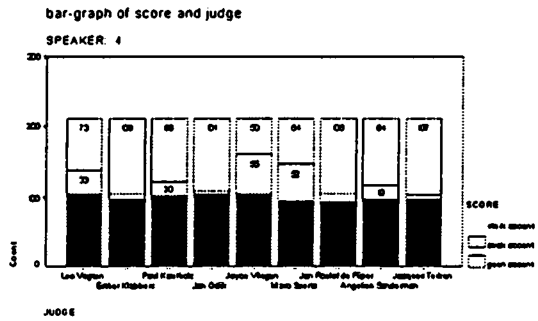
```
/*  van 1 spreker.
/* 'differ' is het berekende verschil van 'totpros' en 'totscore'.
/* 'text'en 'sent' zijn meegenomen omdat dat gemakkelijker zoekt in de
/* tekst.

 USE ALL.
 SORT CASES BY speaker .
 SPLIT FILE BY speaker .
 TEMPORARY .
 SELECT IF ((differ LE -1.33) OR (differ GE 1.33)) .
 LIST text sent wordnr totpros totscore differ aantal0 aantal1 aantal2 .
 EXECUTE .
```

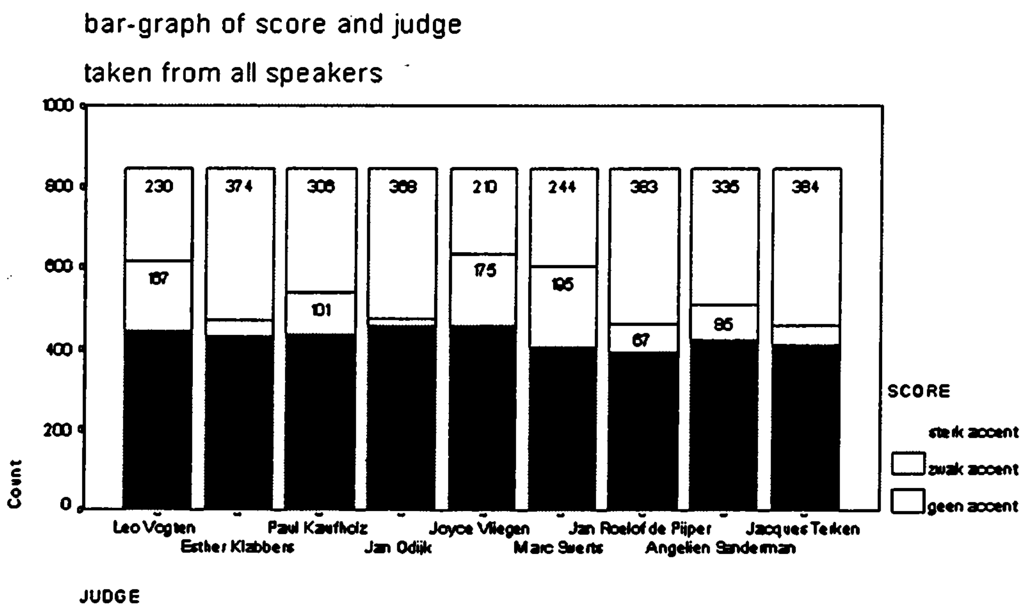
B: Bar-charts.


bar-graph of score and judge
SPEAKER: 1


bar-graph of score and judge
SPEAKER: 2


bar-graph of score and judge
SPEAKER: 3


bar-graph of score and judge
SPEAKER: 4


bar-graph of score and judge
taken from all speakers
1000
800
600
400
200
0
Count
230
374
306
368
210
244
383
335
384
167
101
175
195
67
85
SCORE
sterk accent
zwak accent
geen accent
Leo Vogten
Esther Klabbers
Paul Kaufholz
Jan Odijk
Joyce Vliegen
Marc Swerts
Jan Roelof de Pijper
Angelien Sanderman
Jacques Terken
JUDGE

C: Table of differences.

In this table empty cells represent a value of 'differ' between -1.33 and 1.33, that is only words on which the difference is less then or equal to -1.33 or greater than or equal to 1.33 are present.

wordnr	speaker1	speaker2	speaker3	speaker4	word
7				1.56	eindigde
16	-1.56				toeschouwers
17				1.33	bezochten
20				1.33	scoorde
31	-1.33				leidde
33	1.78		1.78	1.78	duel
35	2.00	2.00	2.00	2.00	Vliet
38	-1.78	-1.78	-1.78	-1.56	rode
49		1.33	1.56	1.56	kaart
53		1.33		1.33	eindigde
59	1.33				duizend
61			1.33	1.56	bezochten
69	-1.78				minuten
76				1.56	Van
82	1.78	1.56		2.00	minuut
94	1.56	1.56		1.78	wedstrijd
109		1.33		1.78	eindigde
117	1.56	1.33	1.33	2.00	bezochten
119	1.56	2.00	1.78	2.00	Gemeentelijk
141	1.33		1.78	2.00	minuut
155	-1.33				PSV
161	1.78		1.78	1.33	minuut
165	-2.00			-1.78	rode
169	1.78	2.00	1.56	2.00	PSV
172	-1.78	-1.78	-1.56	-1.78	man
175	1.78	1.78	2.00	1.78	Vlak
179				1.56	bepaalde
181	1.56	1.78	1.33	1.78	PSV
194		1.33	2.00	1.78	duel

D: The complete collection of texts with Goalgetters prosodic annotation.

Tekst 1:
1.1 het "duel tussen "Vitesse en "Ajax / eindigde in "@nul // - "@een///
1.2 "negenentachtig honderd "zesentwintig "toeschouwers / bezochten "Monnikenhuizen///
1.3 "Litmanen scoorde in de "drieentachtigste "minuut / voor "Ajax///
1.4 "scheidsrechter van "Vliet / "leidde het duel///
*1.5 van Vliet deelde "geen "rode "kaarten uit///
1.6 "Reuser en "Schulp van "Ajax / kregen een "gele kaart///

Tekst 2:
2.1 "Feyenoord tegen "Sparta / eindigde in "@een // - "@een///
2.2 "vijfentwintig duizend "toeschouwers / bezochten de "Kuip///
2.3 "Sparta nam na "achtendertig "minuten de "leiding / door een "doelpunt van van der "Laan///
2.4 in de "zeventigste minuut / bepaalde "Koeman de "eindstand / op "@een // - "@een///
2.5 "scheidsrechter "Luinge / "leidde de wedstrijd///
2.6 van "Bronckhorst van "Feyenoord / kreeg een "gele "kaart///

Tekst 3:
3.1 het "duel tussen Roda "JC en "PSV / eindigde in "@twee // - "@twee///
*3.2 "vijfenveertig honderd "bezoekers / bezochten het Gemeentelijk "Sportpark///
3.3 Roda "JC nam in de "drieenzestigste "minuut de "leiding / door een "doelpunt van de % "aanvaller "Babangida///
3.4 in de "negenenzestigste minuut / tikte "Valckx de "bal in het "verkeerde "doel///
3.5 "twee minuten "later / scoorde de % "PSV speler "Vink///
*3.6 in de "tachtigste minuut / kreeg "Cocu een "rode "kaart // dus moest PSV met "tien "man "verder spelen///
*3.7 vlak voor het "eindsignaal / bepaalde de % PSV speler "Jonk de "eindstand / op "@twee // - "@twee///
3.8 "scheidsrechter Van "Hulten / "leidde het duel///
3.9 "Klomp van "PSV kreeg een "gele "kaart///
3.10 van de "Luer van Roda "JC / kreeg "ook een gele kaart///

Tekst 4:
4.1 "Sparta nam in de "achtste "minuut de "leiding / door een "treffer van de % "verdediger de "Bruin///
4.2 in de "zevenenveertigste minuut / scoorde van der "Laan van Sparta///
4.3 "negenentwintig minuten "later / kreeg "Hansma een "rode "kaart // dus moest Sparta met "tien "*man* "verder spelen///
4.4 in de "tachtigste minuut / bepaalde de % "aanvaller "Fortes de "eindstand / op "@vier // - "@twee///
4.5 "scheidsrechter van "Dijk / "leidde het duel///
*4.6 van Dijk deelde "geen "rode "kaarten uit///
4.7 hij deelde "vijf "gele kaarten uit///