GoalGetter: Predicting Contrastive Accent in Data-to-Speech Generation

Mariët Theune*

Abstract

This paper addresses the problem of predicting contrastive accent in spoken language generation. The common strategy of accenting 'new' and deaccenting 'old' information is not sufficient to achieve correct accentuation; generation of contrastive accent is required as well. I will discuss a few approaches to the prediction of contrastive accent, and propose a practical solution which avoids the problems these approaches are faced with. These issues are discussed in the context of GoalGetter, a data-to-speech system which generates spoken reports of football matches on the basis of tabular information.

Introduction

In language generation systems which produce spoken output, it is important to produce a natural sounding accentuation pattern for each generated sentence. Unnatural sounding speech output is unpleasant to listen to and may be difficult to understand. However, the accentuation pattern should not only be natural sounding but it should also be appropriate with respect to the meaning of the sentence.

In spoken language, accent placement has a major influence on interpretation. Sentences having the same surface structure but a different accentuation pattern may express very different meanings. A well-known example is the sentence *Mary only introduced Bill to Sue* (Rooth (1992)), which can have, among others, the following two accentuation patterns (accented words are given in italics)[1]:

(1) a Mary only introduced *Bill* to Sue
b Mary only introduced Bill to *Sue*

The accentuation patterns presented above each give rise to a different interpretation of the sentence. The accentuation pattern in (1)a indicates that Mary introduced only one person to Sue, and that person was Bill, whereas (1)b conveys that Mary introduced Bill to only one person, and that was Sue.

*This research was carried out at IPO, Center for Research on User-System Interaction, within the framework of the Priority Programme Language and Speech Technology (TST). The TST-programme is sponsored by NWO (the Netherlands Organization for Scientific Research).

[1]For the sake of clarity, in this and the following examples only relevant words are marked for accentuation; e.g., in (1)a/b it is irrelevant whether *Mary* is accented or not and therefore no accentuation is indicated for this word.

When automatically generating spoken output, it is essential that the accentuation pattern assigned to each sentence is in accordance with the intended meaning. If the example sentence *Mary only introduced Bill to Sue* should be interpreted as $\exists!$**x[introduce(mary,x,sue)] & introduce(mary,bill,sue)**, for instance serving as an answer to the question *Who did Mary introduce to Sue?* pronouncing it as in (1)b would be inappropriate and cause the hearer to be confused. The hearer would be faced with conflicting information: the context of the utterance (= the preceding question) would suggest the interpretation given above, whereas the accentuation pattern would give rise to the interpretation $\exists!$**x[introduce(mary,bill,x)] & introduce(mary,bill,sue)**.

It will be clear that, ideally, a spoken language generation system should always assign a correct accentuation pattern to the sentences it generates. In many cases, more than one accentuation pattern can be said to be 'correct', i.e., be in accordance with the intended meaning. What counts as a correct accentuation pattern depends on many factors, including the syntactic and semantic features of the output sentence and its relation to the discourse context. In this paper, I will concentrate on *contrast* as an important discourse semantic factor that must be taken into account for the generation of correct accentuation patterns. I will propose a way of detecting the presence of contrastive information and using this as a basis for the assignment of pitch accent. This will be done within the framework of the GoalGetter system, a data-to-speech system[2] which generates football reports from tabular data.

This paper is structured as follows. After a short introduction to GoalGetter, I will explain the system's original accentuation strategy and explain why this strategy sometimes produced incorrect accentuation patterns (section 1). Since I argue that this could be improved by adding contrastive accent, I will then discuss some existing approaches to contrast and show that these approaches are not attractive as a basis for implementation (section 2). After that, I discuss a practical method for the prediction of contrastive accent which has by now been implemented in GoalGetter, and could be implemented in other data-to-speech systems as well (section 3). In section 4, I discuss some future work. Finally, some conclusions are presented.

1 Accentuation in GoalGetter

Since GoalGetter is described in Klabbers (1997) (this volume), I will only give a very short overview of the system. For further details I refer to Klabbers et al. (1996), Klabbers et al. (1997) and Theune et al. (1997).

The GoalGetter system produces football reports in the form of a spoken monologue in Dutch. These reports are automatically generated on the basis of Teletext pages which contain tabular information on football matches played in the Dutch First Division. The system has two main modules, a language generation module (LGM) and a speech generation module (SGM). The LGM uses the football data from the input Teletext page to generate a written football report, which is an-

[2] Such systems are sometimes called 'concept-to-speech' systems.

notated with prosodic markers, including accentuation markers. This annotated text is input to the SGM, which turns it into a speech signal. Since the assignment of accentuation markers is done in the LGM, I will give a brief description of this module only, restricting the description to those aspects which are relevant for accentuation.

The input for the LGM is a table containing data on a particular football match, which are automatically derived from the information on a Teletext page. This table is converted into an internal data structure which has the form of a record with fields, as shown (partially) in Figure 1.

$$\left[\begin{array}{ll}
\text{teams:} & \left[\begin{array}{ll} \textit{teampair} & \\ \text{home_team:} & \textit{teamtype} \\ \text{visitors:} & \textit{teamtype} \end{array}\right] \\
\text{goallist:} & \left\{\begin{array}{l} \textit{list_of_goalevents} \\ \left[\begin{array}{ll} \textit{goal_event} & \\ \text{team:} & \textit{teamtype} \\ \text{player:} & \textit{playertype} \\ \text{minute:} & \textit{integer} \\ \text{type:} & \textit{goaltype} \end{array}\right] , \ldots \end{array}\right\} \\
\text{result:} & \textit{resulttype} \\
\text{cardlist:} & \textit{list_of_cardevents} \\
\text{referee:} & \textit{refereetype} \\
\text{number_of_spectators:} & \textit{integer}
\end{array}\right]$$

Figure 1: Data structure containing match data

The fields of this record can be expressed by one or more syntactic templates, which are syntactic tree structures containing slots for variable expressions. The filling of the slots depends mainly on conditions on the *Discourse Model*, which contains information about which linguistic expressions have been used in the preceding text, and what they referred to. Rules formulated in terms of this Discourse Model make it possible to use various referential expressions (proper names, pronouns, definite descriptions, etc.) appropriately. When a new sentence has been generated, the Discourse Model is updated accordingly.

The accentuation pattern of each generated sentence is determined on the basis of its syntactic structure and its relation to the preceding text. The accentuation algorithm is based on a version of Focus-Accent Theory (Dirksen (1992), Dirksen and Quené (1993)) and works as follows. First the system determines which parts of the generated sentence are out of focus and should therefore not be accented. This is done on the basis of information in the Discourse Model. Then, partly language-specific accentuation rules determine the distribution of accents, taking both the syntactic structure of the sentence and the focus information into account. Information about these syntax-based rules can be found in Theune et al. (1997). Here I will only discuss the semantic factors which are currently used to determine which phrases are out of focus.

In GoalGetter, a word or phrase will be regarded as being out of focus, and therefore not to be accented, for two reasons: if it is 'unaccentable' or if it conveys 'given' information. To determine if a word is unaccentable, the system simply checks if it belongs to a pre-defined list of words which normally do not receive an accent, e.g., certain function words. The second case is more interesting. As was observed by Halliday (1967), Chafe (1976), Brown (1983) and others, accent can function as a marker of information status: phrases expressing 'new' information are normally accented, while phrases expressing 'given' (or 'old') information are not.

In order to exploit this relationship between accent and information status, the GoalGetter system uses rules to determine whether a certain phrase expresses given information. These rules are based on the theory proposed by van Deemter (1994), who distinguishes two kinds of givenness: object-givenness and concept-givenness. A phrase is regarded as object-given if it refers to a discourse entity that has been referred to earlier in its local discourse domain, which in the present implementation consists of all preceding sentences in the same paragraph. Whether this situation holds can be checked in the Discourse Model. The following fragment can serve as an illustration.[3]

(2) a In the fifth minute, Kluivert scored a goal for Ajax.
b Ten minutes later, the forward had his second goal noted.

In this example, the phrases *the forward* and *his* in (2)b will be regarded as object-given, and therefore deaccented, because they refer to an entity (Kluivert) which was referred to earlier in the same paragraph (i.e., in the preceding sentence). Note that the example shows that object-givenness does not depend on the surface form of the referring expression, but only on its referent.

The second kind of givenness, concept-givenness, occurs if the root of a word has the same denotation as the root of a preceding word in the local discourse domain, or if the concept expressed by the second word subsumes the concept expressed by the first word. Sentence (2)b contains two instances of the first case: the words *minutes* and *goal* are regarded as concept-given due to the presence in the preceding sentence of the words *minute* and *goal* respectively.

Although the strategy of deaccenting given information usually produces correct accentuation patterns, in some cases too many words are deaccented. Using only the given/new distinction as a basis for accentuation may lead to accentuation patterns like the following:

(3) a After three minutes, *Feyenoord* took the lead through a goal by *Koeman.*
b In the sixth minute, *Kluivert* kicked a penalty home for *Ajax.*
c Ten minutes later, *Larsson* scored for Feyenoord.

These three sentences were all generated as part of the same paragraph. In (3)c, the word *Feyenoord* is deaccented due to givenness, because of the previous

[3] Originally, this and the following examples of generated sequences are in Dutch. Since English and Dutch behave in a similar fashion with respect to accentuation I only show the English translations of the original sentences.

mention of *Feyenoord* in (3)a. This wrongly creates the impression that Kluivert scored for Feyenoord, just like Larsson. We see that the generated accentuation pattern does not fit together with the meaning of the sentence. To remedy this, *Feyenoord* should receive contrastive accent, indicating its contrast to *Ajax* in (3)b.

Examples like (3) illustrate what was already suggested by Chafe (1974), and - more recently - by Hirschberg (1992), van Deemter (1994) and Prevost (1995), namely that the given/new distinction is not sufficient to make predictions about accent: it is also necessary to distinguish contrastive accent. In order to generate the correct accentuation patterns for sentences like (3)c, the accentuation rules of GoalGetter should therefore be augmented with an algorithm for the assignment of contrastive accent. This means that the system must be able to recognize contrastive information, which is not a trivial problem. Before I describe the practical solution I implemented in GoalGetter, I will first discuss some theories on the prediction of contrastive accent.

2 Approaches to contrastive accent

In this section I will give a short and informal overview of three different approaches to the prediction of contrast, and point out their disadvantages. The discussion will be restricted to examples involving two subsequent sentences. The three approaches to contrast that I will discuss were proposed by Prevost (1995), van Deemter (1995) and Pulman (1997). They make use of alternative sets, parallelism and contrariety, and higher order unification respectively.

The theory of contrast proposed by Prevost (1995) was inspired by the 'alternative semantics' of Rooth (1992).[4] In Prevost's approach, an item receives contrastive accent if it co-occurs with another item that belongs to its 'set of alternatives', i.e., a set of different items of the same type. Prevost actually implemented his theory in a small generator, which can produce the responses in discourses like the following:

(4) Q: I know the American amplifier produces muddy treble,
but what kind of treble does the British amplifier produce?
A: The *British* amplifier produces *clean* treble

In the example, the two amplifiers are in each other's alternative sets, and so are the two kinds of treble. Because of the presence in the question of *American* and *muddy*, in the answer contrastive accent is assigned to *British* and *clean*.

There are two main problems with this approach. First, as Prevost himself notes, it is difficult to define exactly which items count as being of 'the same type'. If the definition is too strict, not all cases of contrast will be accounted for. On the other hand, if it is too broad, then anything will be predicted to contrast with anything. Prevost gives the following problematic example:

(5) While *he* intently watched the *clock*, *she* watched the *game*.

[4] Although Rooth deals with contrastive accent as well, I will not discuss his theory because it is purely aimed at the *interpretation* of focus (including contrastive accent), not its prediction.

This is a clear case of contrast, but it does not seem appropriate to regard *clock* and *game* as alternatives of each other, since they do not obviously share the same type. Allowing them to count as alternatives would mean an unwanted broadening of the notion of 'alternative set'.

A second problem is that there are cases where there is a clear co-occurrence of items of the same type, but no contrast, as in the following example from the football domain:

(6) a After three minutes, *Feyenoord* took the lead through a goal by *Koeman*.
b This caused *Ajax* to fall behind.
c Ten minutes later *Larsson* scored for Feyenoord.

Prevost's theory would predict *Feyenoord* in (6)c to have a contrastive accent, because the two teams Ajax and Feyenoord are obviously in each other's alternative set. In fact, though, *Feyenoord* should be normally deaccented due to givenness. This shows that the presence of an alternative item does not always trigger contrastive accent.

In the approach proposed by van Deemter (1995), contrast is accounted for in terms of parallelism and contrariety. The cases of contrast discussed above can be easily explained through a notion of parallelism which is closely linked to syntax (see, for instance, the proposal in Prüst (1992)). Both (4) and (5) show a clear parallelism between the succeeding sentences or clauses, while the absence of contrastive accent on *Feyenoord* in (6)c can be explained through a lack of parallelism between (6)b and (6)c.

Still, there are many examples of contrast which seem to lack parallelism. Van Deemter uses the notion of contrariety to account for these cases. Informally defined, two sentences (or clauses) are contrary to each other if they cannot be true at the same time. If two sentences contain two items which are 'contrastible' and whose substitution by the same constant will cause the sentences to be contrary to each other, then these sentences are said to stand in a contrast relationship and the contrastible items will receive contrastive accent. Inequality of denotations is the only condition determining whether two items are contrastible.

Van Deemter gives (7) as an example. If we assume that being an organ mechanic implies knowing much about organs, as stated in the meaning postulate (8), then replacing Mozart by Bach will result in a contrariety. This correctly predicts a contrastive accent on Bach and Mozart.

(7) *Bach* was an organ mechanic; *Mozart* knew little about organs
'Bach was an organ mechanic; Bach knew little about organs'

(8) $\forall x[\mathrm{organ\_mechanic}(x)] \Rightarrow [\mathrm{know\_much\_about\_organs}(x)]$

According to van Deemter, contrastive accent will also fall on those items which, after replacing them by the same constant, cause two sentences to be logically equivalent, as shown in (9).

(9) *Seven* is a prime number and so is *thirteen*
'Seven is a prime number and so is seven'

Apart from the fact that it is not immediately clear how this approach could be implemented in a generation system - checking for contrarieties would certainly require an impossible amount of world knowledge - there is a more important problem with this theory. Van Deemter's condition for 'contrastible items' is extremely permissive, allowing him to avoid the problems which Prevost encounters with examples like (5).[5] However, this liberal notion of contrastibility forces van Deemter to use a severe restriction on what counts as contrast: contrarieties (or equivalences) which are reached through more than one substitution do not qualify for contrastive stress, because otherwise far too many cases of contrastive stress would be predicted. Any pair of sentences of the form (NP_1 VP_1), (NP_2 Negation VP_2) or (NP_1 VP_1), (NP_2 VP_2) would then always count as contrastive, since substitution by the same constant of the NP's and of the VP's at the same time would lead to a contrariety or equivalence.

However, many examples of contrastive accentuation can only be explained if at least two pairs of items are substituted, because substitution of only one pair does not lead to a contrariety or equivalence. These cases cannot be accounted for by the theory. An example from the football domain is (10), where an equivalence (cf. (11)) can only be reached if the pairs *Koeman* - *Kluivert* and *fifth* - *twelfth* are substituted by a constant.

(10) In the *fifth* minute, the referee handed *Koeman* a yellow card; *Kluivert* received a yellow card in the *twelfth* minute
'In the fifth minute, the referee handed Koeman a yellow card; Koeman received a yellow card in the fifth minute'

(11) $\forall x[\text{referee\_hand\_card\_to}(x)] \Leftrightarrow [\text{receive\_card}(x)]$

The examples (7) and (10) can both be explained by Prevost's alternative set theory.

Another approach to the generation of contrastive accent is advocated by Pulman (1997), who proposes to use higher order unification (HOU) for the interpretation and prediction of focus, including contrastive accent. (See also Gardent and Kohlhase (1996) and Gardent et al. (1996).) Pulman makes use of equivalences like the following, which can be used for both interpretation and prediction of focus, and which operate at the level of quasi-logical forms or QLFs (Alshawi and Crouch (1992)):

(12) assert(F,S) $\Leftrightarrow$ S
if
B(F) = S
& context(C)
& P(A) = C
& parallel(B $\bullet$ F, P $\bullet$ A)

[5]Although this particular example could be explained through parallelism in van Deemter's theory, there are other similar examples which do not show parallelism, e.g., 'While the *clock* was all *he* was paying attention to, *she* was watching the *game*'

This says that the QLF S with focus on F is equivalent to S if there is some sentence in the context with a QLF C, where C contains an item A that is parallel to F, while the background P of C (i.e., C after abstracting over A) is parallel to background B of S. Pulman does not define exactly when two items are parallel. Using HOU, equivalence (12) can be resolved in order to predict the landing place of focus markers in a generated sentence, with S being the QLF of this sentence. Pulman illustrates this with the following example, which I have simplified somewhat.

In the context of a system which generates information about the operation of some machinery, a user might ask *Do I put the card into the slot?*, which would be analysed as (13). Assuming that the correct operation at this point actually is that you put a disc into the slot, the semantics of the response by the system will be as given in (14).

(13) $\exists$xy[put(user,x,y) & card(x) & slot(y)]

(14) $\exists$xy[put(user,x,y) & disc(x) & slot(y)]

Now the equivalence in (12) will be resolved as follows. S is the QLF of the sentence to be generated, represented in (14). C will be equated to (13), where P = B = λ**P**$\exists$**xy[put(user,x,y) & P(x) & slot(y)]**, A = **card** and F = **disc**. This means that the surface expression generated for **disc** should be marked for focus (in this case, contrastive accent).

Like van Deemter (1995), Pulman makes crucial use of parallelism, a notion which is as difficult to define as Prevost's alternative set. Pulman does not give a full definition of which items count as being parallel, but states that "to be parallel, two items need to be at least of the same type and have the same sortal properties" (Pulman (1997), p. 90). This condition is rather similar to Prevost's conditions on alternative sets. Consequently, Pulman's theory faces the same problem as Prevost's, namely that of defining when two items are of the same type. Like Prevost, Pulman can only explain the contrast in example (5) if *clock* and *game* count as being parallel, something which is not obvious.

Pulman has a theoretical advantage over Prevost in that he stresses that two sentences should not only contain some parallel items to warrant contrastive stress, as in Prevost (1995), but that the 'background' parts of the sentences should be parallel as well. In principle, this more restrictive condition on contrastive accent makes it possible for Pulman to account for examples like (6), which Prevost cannot explain: presumably, Pulman would not regard the backgrounds of (6)b and c as parallel.[6] However, as long as Pulman does not give a proper definition of parallelism, it is impossible to say what his theory will or won't predict.

As Gardent et al. (1996) point out, a HOU approach can take world knowledge into account when solving equations as in the example given above. They do not give an explicit description of *how* world knowledge can be used in solving equivalences, but presumably it could be done by making use of meaning postulates like those in (8) and (11) to solve those cases where the semantic representations

[6]Prevost (personal communication) claims that he also looks for semantic parallelism between sentences, but this is not apparent from Prevost (1995).

of two sentences do not unify. For example, the contrast between the two clauses of (10) can be predicted if C in (12) is not equated to the direct semantic representation of the first clause, but to its equivalent according to (11). In this way (assuming a proper definition of parallelism is available), Pulman should be able to make the correct predictions for both (7) and (10). A similar solution might be possible for van Deemter (1995): by taking entailments and equivalences into account for the determination of parallelism, examples like (10), but also (7) could be accounted for in terms of parallelism. This way, checking for contrariety might become unnecessary.

To conclude, we have seen that a notion of semantic parallelism in combination with world knowledge seems to make the best predictions of contrast. However, a good definition of parallelism is lacking, and the encoding of world knowledge is a notorious problem. Even in a small domain like football reports the explicit enumeration of all possible semantic entailments and equivalences seems hardly feasible. Fortunately, data-to-speech systems like GoalGetter, the input of which is formed by typed and structured data, offer a simple way of automatically establishing semantic parallelism, with no need to explicitly encode world knowledge. In the next section, I will discuss how this can be done.

3 Contrastive accent in a data-to-speech system

The method I propose, and which has been successfully implemented in GoalGetter, is based on the simple principle that two sentences which express the same type of data structure (and therefore express similar information) should be regarded as contrastive. Contrastive accent should be assigned to those parts of the second sentence that express values which differ from those in the data structure expressed by the first sentence.

The idea behind this is the following. As we saw in the preceding section, for establishing contrast it is not sufficient to directly compare the semantic representations of two sentences: we need to use world knowledge to establish whether the sentences are semantically parallel, i.e. whether they describe similar situations or events. In our system this 'real world' information is readily available in the form of the data structures that are expressed by the sentences. We may consider two sentences semantically parallel if they express information contained in data structures of the same type, without caring about the specific linguistic forms chosen to convey this information. In this way, we can avoid the problems encountered by most of the theories discussed in section 2, as I will show in the rest of this section.

I will use example (3) from section 1 as an illustration. As was explained in that section, GoalGetter's football reports are generated on the basis of a typed data structure which is derived from the information on a Teletext page. The field `goallist` of this data structure contains a sequence of records of type *goal_event*, each record specifying the team for which a goal was scored, the player who scored, the time and the kind of goal: normal, own goal or a goal resulting from a penalty. The last two sentences of example (3) both express such a *goal_event* data structure, given in Figure 2, so they are regarded as contrastive, even though they show no direct syntactic or semantic parallelism.

$$\text{goal\_event (3)b} \quad \begin{bmatrix} \text{team:} & \text{Ajax} \\ \text{player:} & \text{Kluivert} \\ \text{minute:} & 6 \\ \text{goaltype:} & \text{penalty} \end{bmatrix}$$

$$\text{goal\_event (3)c} \quad \begin{bmatrix} \text{team:} & \text{Feyenoord} \\ \text{player:} & \text{Larsson} \\ \text{minute:} & 16 \\ \text{goaltype:} & \text{normal} \end{bmatrix}$$

Figure 2: Data structures expressed by (3)b and (3)c.

As can be seen in Figure 2, all the fields of the *goal_event* record expressed by (3)c have different values from that of (3)b. This means that all phrases in (3)c expressing the values of those fields should receive contrastive accent, including *Feyenoord*, despite its givenness. Note that the value of the `goaltype` field is not expressed in the surface structure of (3)c; however, if it were, it would receive a contrastive accent (e.g., *Ten minutes later, Larsson scored a* normal *goal for Feyenoord*).

Another example where lack of contrastive accent in GoalGetter used to lead to an incorrect accentuation pattern is the following sequence. Using only the given/new distinction without contrastive accent would lead to the following accentuation pattern:

(15) a In the sixteenth minute, the *Ajax* player *Kluivert* kicked the ball into the wrong goal.

b Twenty minutes later, *Overmars* scored for Ajax.

The deaccentuation of *Ajax* in (15)b gives the impression that both Kluivert and Overmars scored for Ajax, while in fact Kluivert scored for the other team through an own goal. Therefore, the second occurrence of Ajax should receive a contrastive accent despite its being given. In the theory of Prevost, this cannot be explained: (15)a does not contain a member of the alternative set of Ajax, so no contrast is predicted. Van Deemter's theory does not predict contrastive accent either, because (15)a and b do not show any parallelism, and contrariety only occurs after substitution of *two* pairs of items, the players and the times. Using Pulman's approach, contrast can only be predicted if the system contains the world knowledge that scoring an own goal means scoring for the opposing team.

The method proposed here does not require additional world knowledge to determine the presence of contrast in (15)b; the contrast can be immediately derived from the data structures expressed by sentences (15)a and b, which are given in Figure 3. A simple comparison of the `team` fields of (15)a and b shows that they have contrasting values, and that the phrase expressing the `team` field in (15)b should receive contrast accent, even though the corresponding value of the previous sentence was not overtly expressed.

goal_event (15)a

team:	Feyenoord
player:	Kluivert
minute:	16
goaltype:	own

goal_event (15)b

team:	Ajax
player:	Overmars
minute:	36
goaltype:	normal

Figure 3: Data structures expressed by (15)a and (15)b.

As we see, one of the advantages of the approach sketched above is that it requires no explicit listing of semantic equivalences or entailments. Only the information (data) which is expressed by a sentence is taken into account for the detection of contrast; which surface form is chosen to express certain information is not important. The discussion of examples (3) and (15) has shown that data can be expressed in an indirect way without influencing the prediction of contrast for the following sentence.

The approach sketched above will also give the desired result for example (6): sentence (6)c will not be regarded as contrastive with (6)b, since (6)c expresses a *goal_event* but (6)b does not. Therefore no contrastive accent will be assigned to *Feyenoord* in (6)c.

The approach can be extended to deal with deaccenting as well. Those parts of a sentence that express values which are identical to values in the data structure from which the previous sentence was generated, should be deaccented. This way, we can account for cases of deaccenting that cannot be handled by GoalGetter's current defocusing strategy, described in section 1. This can be illustrated by example (16), a variant of (10). The corresponding data structures are given in Figure 4. These structures are of type *card_event*, and describe at which time which player received a card of which colour.

(16) a In the *fifth* minute, *Koeman* was sent off the field.
b *Kluivert* received a red card in the *twelfth* minute.

card event (16)a

player:	Koeman
minute:	5
cardtype:	red

card event (16)b

player:	Kluivert
minute:	12
cardtype:	red

Figure 4: Data structures expessed by (16)a and (16)b.

Sentence (16)a expresses its underlying data in an implicit manner, leaving the colour of the card unspecified but inferrable. Sentence (16)b does explicitly mention the colour. Because the kind of card in this sentence is the same as in (16)a, the phrase expressing it (*red card*) is deaccented. This is not predicted by the defocusing strategy described in section 1, since in (16)a the type of card is not explicitly mentioned, and is therefore not detected by the defocusing algorithm. However, by looking at the data structures of (16)a and b, we can see that the values of the card feature are identical. The phrase *red card* in (16)b should therefore be deaccented. The result is the correct accentuation pattern as shown in (16), which will confirm the inference of the hearer that Koeman was shown a red card too.

Obviously, the proposed method places a great responsibility on the data structures that are used. The problem of defining parallelism is shifted to the design of the data structures: they must be set up in such a way that parallel items get assigned identical data types. It is still an open question whether it would be possible to specify general conditions on data structures, which they should meet in order to be usable for establishing contrast. So far, it seems that any data structure which is a plausible representation of the relevant domain, and which is rich enough to reflect the relations between objects in this domain, should be usable. This is confirmed by the fact that the data structure of GoalGetter was not designed for the prediction of contrast, but still proved to be suitable for this purpose.

4 Future work

The next step will be to see if the method described in the previous section can also be applied in another system, namely the OVIS system which is currently being developed in the Priority Programme Language and Speech Technology of NWO, the Netherlands Organization for Scientific Research. The OVIS dialogue system will provide information about public transport in the Netherlands. There already exists a typed data structure for this system, which has been designed independently from language generation. If this structure turns out to be usable for deriving contrast relations, this will prove that the applicability of the proposed method is not limited to GoalGetter.

Additionally, the principle on which the proposed method is based has to be further refined. For example, an open question which still remains is at which level data structures should be compared. Figures 3 and 4 presented data structures of type *goal_event* and *card_event* respectively. Since these data structures are of different types, currently they are not predicted to be contrastive. However, both are subtypes of a more general *event* type, which has only the fields `team`, `player`, and `minute`. For this reason, *goal_event* and *card_event* might have to be considered as contrastive after all. Examples like (17) seem to point in this direction.

(17) a In the eleventh minute, Ajax took the lead through a goal by *Kluivert*.
b Shortly after the break, the referee handed *Koeman* a yellow card.
c Ten minutes later, *Kluivert* scored for the second time.

The fact that Kluivert can be accented in (17)c can only be explained if (17)c is potentially contrastive to (17)b; otherwise, the second mention of Kluivert would be deaccented due to givenness, like *Feyenoord* in (6)c.

How such cases should be dealt with, will be the subject of further research. In general, the possibility of contrast between types and their subtypes (not only of events, but also of objects) should be further investigated. Presumably, both domain and discourse context play an important role here.

5 Conclusions

In this paper I have shown how the strategy of deaccenting given information can lead to incorrect accentuation patterns if contrast is not taken into account. Contrastive information should receive an accent, even if it is given. Approaches to the prediction of contrast which have been proposed in the literature are not attractive as a basis for implementation. The approach proposed by Prevost (1995) does not take parallelism between sentences into account and therefore does not always make the correct predictions. The contrast theory of van Deemter (1995) is too restrictive and cannnot account for all cases. Pulman (1997) does not give a proper definition of parallelism, and like the theory of van Deemter (1995), it requires a large amount of world knowledge in order to make the right predictions. Since it would be impossible to encode all relevant world knowledge, another solution must be found.

As an alternative, I have proposed a practical method to the assignment of contrastive accent in data-to-speech systems. In contrast to the approaches advocated by Prevost, van Deemter and Pulman, this method does not require a universal definition of alternative or parallel items. Also, the fact that determination of contrast is based on the information content of sentences obviates the need for explicitly encoding world knowledge; we can make use of the world knowledge which is already incorporated in the design of the data structures that are to be expressed. The use of these data structures for the prediction of contrastive accent is based on a general principle, which should be applicable in any system that generates sentences from a typed data structure.

The proposed approach has been implemented in the GoalGetter system and will be implemented in the OVIS system in the near future.

References

Alshawi, H. and R. Crouch (1992). Monotonic semantic interpretation. In *Proceedings of the 30th Annual Meeting of the ACL*, pp. 32–39.

Brown, G. (1983). Prosodic structure and the given/new distinction. In D. R. Ladd and A. Cutler (Eds.), *Prosody: Models and Measurements*, pp. 67–77. Berlin: Springer Verlag.

Chafe, W. (1974). Language and consciousness. *Language 50*, 111–133.

Chafe, W. (1976). Givenness, contrastiveness, definiteness, subjects, topics and points of view. In C. N. Li (Ed.), *Subject and Topic*, pp. 25–55. New York:

Academic Press.

Dirksen, A. (1992). Accenting and deaccenting: a declarative approach. In *Proceedings of COLING 1992, Nantes, France*, pp. 865–869. IPO MS. 867.

Dirksen, A. and H. Quené (1993). Prosodic analysis: the next generation. In van Heuven and Pols (Eds.), *Analysis and Synthesis of Speech: Strategic Research Towards High-Quality Text-to-Speech Generation*, pp. 131–144. Berlin - New York: Mouton de Gruyter.

Gardent, C. and M. Kohlhase (1996). Focus and higher-order unification. In *Proceedings of COLING 1996*, pp. 430–435.

Gardent, C., M. Kohlhase, and N. van Leusen (1996). Corrections and higher-order unification. In *Proceedings of KONVENS, Bielefeld*, pp. 268–279.

Halliday, M. (1967). Notes on transitivity and theme in English. *Journal of linguistics 3*, 199–244.

Hirschberg, J. (1992). Using discourse context to guide pitch accent decisions in synthetic speech. In G. Bailly, C. Benoît, and T. Sawallis (Eds.), *Talking Machines: Theories, Models and Designs*, pp. 367–376. Elsevier Science Publishers B.V.

Klabbers, E. (1997). Speech output generation in GoalGetter. In K. van Deemter, J. Landsbergen, J. Odijk, and G. Veldhuijzen van Zanten (Eds.), *CLIN VII, papers from the seventh CLIN meeting.*

Klabbers, E., J. Odijk, J. de Pijper, and M. Theune (1996). GoalGetter: From Teletext to speech. *IPO Annual Progress Report 31*, 61–75.

Klabbers, E., J. Odijk, J. de Pijper, and M. Theune (1997). From data to speech: a generic approach. IPO MS 1202.

Prevost, S. (1995). *A Semantics of Contrast and Information Structure for Specifying Intonation in Spoken Language Generation.* Ph. D. thesis, University of Pennsylvania.

Prüst, H. (1992). *On discourse structuring, VP anaphora and gapping.* Ph. D. thesis, University of Amsterdam.

Pulman, S. (1997). Higher order unification and the interpretation of focus. *Linguistics and Philosophy 20*, 73–115.

Rooth, M. (1992). A theory of focus interpretation. *Natural Language Semantics 1*, 75–116.

Theune, M., E. Klabbers, J. Odijk, and J. de Pijper (1997). Computing prosodic properties in a data-to-speech system. In *Proceedings of the Workshop on Concept-toSpeech Generation Systems, ACL/EACL 1997*, Madrid, pp. 39–45.

van Deemter, K. (1994). What's new? A semantic perspective on sentence accent. *Journal of Semantics 11*, 1–31.

van Deemter, K. (1995). Contrastive stress, contrariety and focus. In P. Bosch and R. v.d. Sandt (Eds.), *Focus & Natural Language Processing.* Cambridge: Cambridge University Press.