

Contrastive accent in a data-to-speech system

Mariët Theune

IPO, Center for Research on User-System Interaction

P.O. Box 513

5600 MB Eindhoven

The Netherlands

theune@ipo.tue.nl

Abstract

Being able to predict the placement of contrastive accent is essential for the assignment of correct accentuation patterns in spoken language generation. I discuss two approaches to the generation of contrastive accent and propose an alternative method that is feasible and computationally attractive in data-to-speech systems.

1 Motivation

The placement of pitch accent plays an important role in the interpretation of spoken messages. Utterances having the same surface structure but a different accentuation pattern may express very different meanings. A generation system for spoken language should therefore be able to produce appropriate accentuation patterns for its output messages.

One of the factors determining accentuation is contrast. Its importance can be illustrated with an example from GoalGetter, a data-to-speech system which generates spoken soccer reports in Dutch (Klabbers et al., 1997). The input of the system is a typed data structure containing data on a soccer match. So-called syntactic templates (van Deemter and Odijk, 1995) are used to express parts of this data structure. In GoalGetter, only ‘new’ information is accented; ‘given’ (‘old’) information is not (Chafe, 1976), (Brown, 1983), (Hirschberg, 1992). However, this strategy does not always lead to a correct accentuation pattern if contrastive information is not taken into account, as shown in example (1).¹

- (1) a In the 16th minute, the Ajax player Kluivert kicked the ball into the wrong goal.
b Ten minutes later, *Wooter* scored for Ajax.

¹All GoalGetter examples are translated from Dutch. Accented words are given in italics; deaccented words are underlined. This is only done where relevant.

The word *Ajax* in (1)b is not accented by the system, because it is mentioned for the second time and therefore regarded as ‘given’. However, this lack of accent creates the impression that Kluivert scored for Ajax too, whereas in fact he scored for the opposing team through an own goal. This undesirable effect could be avoided by accenting the second occurrence of *Ajax* in spite of its givenness, to indicate that it constitutes contrastive information.

2 Predicting contrastive accent

In this section I discuss two approaches to predicting contrastive accent, which were put forward by Scott Prevost (1995) and Stephen Pulman (1997).

In the theory of contrast proposed in (Prevost, 1995), 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. There are two main problems with this approach. First, as Prevost himself notes, it is very 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. A second problem is that there are cases where co-occurrence of two items of the same type does not trigger contrast, as in the following soccer example:

- (2) a After six minutes *Nilis* scored a goal for *PSV*.
b This caused *Ajax* to fall behind.
c Twenty minutes later *Cocu* scored for PSV.

According to Prevost’s theory, *PSV* in (2)c should have a contrastive accent, because the two teams Ajax and PSV are obviously in each other’s alternative set. In fact, though, there is no contrast and *PSV* should be normally deaccented due to givenness. This shows that the presence of an alternative item is not sufficient to trigger contrast accent.

Another approach to contrastive accent is advocated by Pulman (1997), who proposes to use higher order unification (HOU) for both interpretation and prediction of focus. Described informally, Pulman’s focus assignment algorithm takes the semantic representation of a sentence which has just been generated, looks in the context for another sentence representation containing parallel items, and abstracts over these items in both representations. If the resulting representations are unifiable, the two sentences stand in a contrast relation and the parallel elements from the most recent one receive a pitch accent (or another focus marker).

Pulman does not give a full definition of parallelism, 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 is rather similar to Prevost’s conditions on alternative sets. Consequently, Pulman’s theory also faces the problem of determining when two items are of the same type. Still, contrary to Prevost, Pulman can explain the lack of contrast accent in (2)c, because obviously the representations of sentences (2)b and (2)c will not unify.

Another advantage, pointed out in (Gardent et al., 1996), is that a HOU algorithm can take world knowledge into account, which is sometimes necessary for determining contrast. For instance, the contrast in (1) is based on the knowledge that kicking the ball into the wrong goal implies scoring a goal for the opposing team. In a HOU approach, the contrast in this example might be predicted by unifying the representation of the second sentence with the entailment of the first. However, such a strategy would require the explicit enumeration of all possible semantic equivalences and entailments in the relevant domain, which seems hardly feasible. Also, implementation of higher order unification can be quite inefficient. This means that although theoretically appealing, the HOU approach to contrastive accent is less attractive from a computational viewpoint.

3 An alternative solution

Fortunately, in data-to-speech systems like GoalGetter, the input of which is formed by typed and structured data, a simple principle can be used for determining contrast. If two subsequent sentences are generated from the same type of data structure they express similar information and should therefore be regarded as potentially contrastive, even if their surface forms are different. Pitch accent should be assigned to those parts of the second sentence that express data which differ from those in the data structure expressed by the first sentence.

Example (1) can be used as illustration. The theory of Prevost will not predict contrastive accent on *Ajax* in (1)b, because (1)a does not contain a member of its alternative set. In Pulman’s approach, the contrast can only be predicted if the system uses the world knowledge that scoring an own goal means scoring for the opposing team. In the approach that I propose, the contrast between (1)a and b can be derived directly from the data structures they express. Figure 1 shows these structures, A and B, which are both of the type *goal_event*: a record with fields specifying the team for which a goal was scored, the player who scored, the time and the kind of goal: normal, own goal or penalty.

A:	<i>goal_event</i>	[<table style="border: none; margin: 0 auto;"> <tr><td>team:</td><td>PSV</td></tr> <tr><td>player:</td><td>Kluivert</td></tr> <tr><td>minute:</td><td>16</td></tr> <tr><td>goaltype:</td><td>own</td></tr> </table>	team:	PSV	player:	Kluivert	minute:	16	goaltype:	own]
team:	PSV											
player:	Kluivert											
minute:	16											
goaltype:	own											
B:	<i>goal_event</i>	[<table style="border: none; margin: 0 auto;"> <tr><td>team:</td><td>Ajax</td></tr> <tr><td>player:</td><td>Wooter</td></tr> <tr><td>minute:</td><td>26</td></tr> <tr><td>goaltype:</td><td>normal</td></tr> </table>	team:	Ajax	player:	Wooter	minute:	26	goaltype:	normal]
team:	Ajax											
player:	Wooter											
minute:	26											
goaltype:	normal											

Figure 1: Data structures expressed by (1)a and b.

Since A and B are of the same type, the values of their fields can be compared, showing which pieces of information are contrastive. Figure 1 shows that all the fields of B have different values from those of A. This means that each phrase in (1)b which expresses the value of one of those fields should receive contrastive accent,² even if the corresponding field value of A was not mentioned in (1)a. This guarantees that in (1)b the proper name *Ajax*, which expresses the value of the **team** field of B, is accented despite the fact that the contrasting team was not explicitly mentioned in (1)a.

The discussion of example (1) shows that in the approach proposed here no world knowledge is needed to determine contrast; it is only necessary to compare the data structures that are expressed by the generated sentences. The fact that the input data structures of the system are organized in such a way that identical data types express semantically parallel information allows us to make use of the world (or domain) knowledge incorporated in the design of these data structures, without having to separately encode this knowledge. This also means

²Sentence (1)b happens not to express the goaltype value of B, but if it did, this phrase should also receive contrastive accent (e.g., ‘Twenty minutes later, *Overmars* scored a *normal* goal’).

that the prediction of contrast does not depend on the linguistic expressions which are chosen to express the input data; the data can be expressed in an indirect way, as in (1)a, without influencing the prediction of contrast.

The approach sketched above will also give the desired result for example (2): sentence (2)c will not be regarded as contrastive with (2)b, since (2)c expresses a goal event but (2)b does not.

4 Future directions

An open question which still remains, is at which level data structures should be compared. In other words, how do we deal with sub- and supertypes? For example, apart from the *goal_event* data type the GoalGetter system also has a *card_event* type, which specifies at what time which player received a card of which color. Since *goal_event* and *card_event* are different types, they are not expected to be contrastible. However, both are subtypes of a more general event type, and if regarded at this higher *event* level, the structures might be considered as contrastible after all. Examples like (3) seem to suggest that this is possible.

- (3) a In the 11th minute, Ajax took the lead through a goal by Kluivert.
 b Shortly after the break, the referee handed Nilis a yellow card.
 c Ten minutes later, *Kluivert* scored for the *second* time.

The fact that it is not inappropriate to accent *Kluivert* in (3)c, shows that (3)c may be regarded as contrastive to (3)b; otherwise, it would be obligatory to deaccent the second mention of *Kluivert* due to givenness, like *PSV* in (2)c. Cases like this might be accounted for by assuming that there can be contrast between fields that are shared by data types having the same supertype. In (3), these would be the **player** and the **minute** fields of structures C and D, shown in Figure 2. This is a tentative solution which requires further research.

C:	<i>card_event</i>	[player:	Nilis]
			minute:	11	
			cardtype:	yellow	
D:	<i>goal_event</i>	[team:	Ajax]
			player:	Kluivert	
			minute:	21	
			goaltype:	normal	

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

5 Conclusion

I have sketched a practical approach to the assignment of contrastive accent in data-to-speech systems, which does not need a universal definition of alternative or parallel items. Because the determination of contrast is based on the data expressed by generated sentences, instead of their syntactic structures or semantic representations, there is no need for separately encoding world knowledge. The proposed approach is domain-specific in that it relies heavily on the data structures that form the input from generation. On the other hand it is based on a general principle, which should be applicable in any system where typed data structures form the input for linguistic generation. In the near future, the proposed approach will be implemented in GoalGetter.

Acknowledgements: This research was carried out within the Priority Programme Language and Speech Technology (TST), sponsored by NWO (the Netherlands Organization for Scientific Research).

References

- Gillian Brown. 1983. Prosodic structure and the given/new distinction. In D.R. Ladd and A. Cutler (Eds.): *Prosody: Models and Measurements*. Springer Verlag, Berlin.
- Wallace Chafe. 1976. Givenness, contrastiveness, definiteness, subjects, topics and points of view. In C.N. Li (Ed): *Subject and Topic*. Academic Press, New York.
- Kees van Deemter and Jan Odijk. 1995. Context modeling and the generation of spoken discourse. Manuscript 1125, IPO, Eindhoven, October 1995. Philips Research Manuscript NL-MS 18 728. To appear in *Speech Communication*, 21 (1/2).
- Claire Gardent, Michael Kohlhase and Noor van Leusen. 1996. Corrections and higher-order unification. To appear in *Proceedings of KONVENS*, Bielefeld.
- Julia Hirschberg. 1992. Using discourse context to guide pitch accent decisions in synthetic speech. In G. Bailly, C. Benoît and T.R. Sawallis (Eds) *Talking Machines: Theories, Models, and Designs*. Elsevier Science Publishers, Amsterdam, The Netherlands.
- Esther Klabbbers, Jan Odijk, Jan Roelof de Pijper and Mariët Theune. 1997. GoalGetter: from Teletext to speech. To appear in *IPO Annual Progress Report 31*. Eindhoven, The Netherlands.
- Scott Prevost. 1995. *A semantics of contrast and information structure for specifying intonation in spoken language generation*. PhD-dissertation, University of Pennsylvania.
- Stephen Pulman. 1997. Higher Order Unification and the interpretation of focus. In *Linguistics and Philosophy* 20.