Developing the prosodic component for Swedish speech synthesis

Horne, Merle; Filipsson, Marcus

DEVELOPING THE PROSODIC COMPONENT FOR SWEDISH SPEECH SYNTHESIS

M. Horne & M. Filipsson
e-mail: Merle.Horne@ling.lu.se
Dept. of Linguistics and Phonetics
University of Lund
Helgonabacken 12
S-223 62 Lund
SWEDEN

ABSTRACT
A proposal is made for the design of an algorithm for determining when pronouns and 'Given' content words should be assigned accents due to a shift in grammatical function.

INTRODUCTION
In a number of recent papers [1-2], we have presented the form of an algorithm for generating prosodic structure for Swedish speech synthesis. One component of this system is a referent-tracker which identifies coreference or identity of sense relationships between lexical or content words. This is a central component since the accent assigning rules are constructed so that words representing 'new' information are assigned focal accents and coreferent content words associated with 'given' infomation are not focally accented. It has been pointed out, however, in a number of places that it is certainly not always the case that only words expressing new lexical information receive accents [3-8]. For example, rhythmical factors can lead to the accenting of a 'given' content word at the beginning of a Prosodic Phrase as in (1) where accented words are written in bold script:

(1) What **TIME** is the **MEETING**?
The **MEETING** is **CANCELLED**.

In (1), the second occurrence of MEETING is accented even though it is contextually given. This accentuating can be explained as due to rhythmic factors, i.e. the preference for an accent or prominence as early as possible in a prosodic phrase [6]. Such accenting of given information can be modelled in a synthesis system, given the possibility of delimiting the prosodic domain for such rhythmically motivated prominences.

In our current algorithm for generating intonation, pronouns, like given content words, are not assigned accents either. Although unaccented pronouns constitute the unmarked case, there are many instances where pronouns are accentuated. They can, for example, be accented due to discourse factors, e.g. in order to signal a shift in attention [9] as in (2):

(2) During the past weeks, the papers have been full of success stories about the talented slalom king and long-time bachelor, Jeff Barkley. He has now announced his engagement to the Spanish beauty Elisa Morales.
SHE is a well-known actress, both in her own country and abroad...

According to the discourse-based thery of 'centering' developed by Grosz et al. [10-11], the accented pronoun SHE, illustrates a local shift establishing *Elisa Morales* as the new 'center' in the discourse segment, where center is regarded as an entity that serves to link an utterance to other utterances in the discourse segment that contains it [11]. That is to say, accenting of pronouns can be used to mark a change in what could be called the most salient local topic. Computationally, such cases of center shifts are difficult to recognize. There is no regular lexical or syntactic structures that correlate with this type of discourse-motivated accentuating of pronouns.

CONTRASTIVE STRESS AND CENTERING
There are, however, situations involving center shifts which are more amenable to computational modelling. Perhaps more common than the ones mentioned above are cases that have been described as involving 'contrastive stress' determined by syntactic parallelism [4, 12-13] in examples like those in (2):

(2) a. Kent hit Bert and then HE hit HIM.
b. Kent hit Bert and then HE slugged HIM

Although the pronouns in (2) refer to contextually given antecedents, they are accented. This has variously been interpreted as functioning to indicate a "shift in semantic roles" [12] or "grammatical functions" [4] in cases where the predicates in the two clauses are identical in sense. In a 'centering' perspective, however, cases of pronominal accentuation like (2) can also be interpreted as involving a local shift of center of attention as regards the relation of the discourse referents to the state or action expressed by the predicate in the two utterances. The change in this relationship is what can be considered to be the 'new' information and is expressed by accentuating the pronouns.

According to Kameyama's work on utterance interpretation [14], an unaccented pronoun normally realizes a *maximally salient entity* in a discourse, whereas an accented pronoun takes the 'complementary preference' of the unstressed counterpart. Salience is here related to two linguistic hierarchies: i) a grammatical function hierarchy [Subj.>Obj.>Obj.2>Others], which reflects the ordering among the 'forward centers', i.e. entities that link an utterance with later utterances in a discourse segment and ii) a nominal expression hierarchy: Unstressed Pronoun>Stressed Pronoun>Definite NP>Indefinite NP, which reflects the expected form of the 'backward center' of an utterance, i.e. the form of a

discourse entity that connects with the most highly ranked forward center in the preceding utterance. According to Kameyama, these hierarchies reflect the fact for example that an unstressed pronoun normally corefers with the matrix Subject in the previous utterance as in (3):

(3) a. John hit Bill. [John>Bill]
   b. and then he hit Tony [he:=John]

In (3b), the preferential interpretation of *he* is the Subject of the preceding utterance, not the Object, *Bill*. On the other hand, if one were to use an accented pronoun in the b) utterance, the most preferred interpretation would be that HE refers to the less salient Object in the a) utterance, *Bill*, as (4) shows:

(4) a. John hit Bill. [John (Subj)>Bill (Obj)]
   b. and then HE hit TONY [(Comp. pref.)
   HE:= Bill (Obj.)]

Although the **interpretation** of the utterances in (4) is computationally possible given information on the accented vs unaccented status of the pronouns, the **generation** of the accentual patterns is not possible without knowledge about the intentions of the actors since the masculine Subject form of the pronoun *he* can potentially refer to both *John* and *Bill*. There is thus no lexical or syntactic information that would allow one to predict whether *he* should be accented or unaccented in (4b).

The accentuation of pronouns due to a corresponding change in grammatical function assigned to them is, nevertheless, possible to model computationally to a certain extent without recourse to speaker intentions if the pronoun reference is unambiguous. This type of pronoun resolution has not, however, been included in existing algorithms for accent assignment, since, as pointed out earlier, the unmarked case is for pronouns to be unaccented. However, failure to account for these marked cases leads to imperfect performance of F0 generating algorithms.

In order to handle cases of accentuated pronouns, one must be able first of all to recognize identity relations between predicates (verbs) in two or more clauses. This is possible in our preprocessor which includes a referent-tracker that identifies coreference relations among content words. What is lacking is an algorithm for pronoun resolution and a procedure for determining whether a pronoun's grammatical function in relation to a given predicate is the same as that of its antecedent if that antecedent also happens to be an argument of the preceding identical predicate.

PRONOUN RESOLUTION AND GRAMMATICAL FUNCTION RECOGNITION IN SWEDISH

According to Fraurud's algorithm for pronoun resolution in Swedish [15], one should first **consider** as a candidate antecedent every NP in the text that: (i) precedes the pronoun in the text, (ii) agrees with the pronoun in number, animacy and gender, (iii) is not coreferential with the subject of the clause in which the

pronoun occurs, (iv) is not in a clause where the pronoun is Subject. One should then **select** as antecedent the NP in the set of candidate antecedents that: (i) is the most recent candidate in the text, i.e. the NP whose head is closer to the pronoun than the head(s) ot the other NP(s), unless: (ii) there is another candidate, which is the Subject of the same clause as the most recent candidate. This last restriction reflects the hierarchy of grammatical functions in Kameyama's analysis.

In order to illustrate how this algorithm for pronoun resolution would apply to Swedish data, we present below a number of examples of texts extracted from literary works where accentuation is indicated by the use of italics. The first example is taken from *Pippi Långstrump* 'Pippi Longstocking' where accenting of the pronoun *HON* 'she' is necessary due to a change in its grammatical function in relation to the predicate *slå sönder* 'break':

(5) Fru Settergren sa: "Inte för att jag precis vill klaga på min Ella, men slå sönder porslin, det gör hon".
Ett rött huvud blev synligt uppe i trappan.
"På tal om Malin", sa Pippi, "så kanske ni undrar om *HON* brukade slå sönder något porslin.".
Mrs. Settergren said : "Not that I want to complain about my Ella, but break porcelain, that she does."
A red head appeared in the stairway. "Speaking of Malin", said Pippi, "maybe you're wondering if *SHE* used to break any porcelain".

Extraction of the identical predicates and associated arguments in the quoted discourse yields the following sequence (where N = New , G = Given); underlining marks the boundary between predicate/argument groupings:

jag	G	det	G	ni	G
klaga	N	gör	G	undrar	N
Ella	N	hon	G	hon	G
slå sönder	N	tal	N	slå sönder	G
porslin	N	Malin	N	porslin	G

Here we see that there are two occurrences of the predicate *slå sönder*. Furthermore, one of the arguments associated with this verb is the pronoun Subject *hon* 'she'. The problem is then to resolve the identity of *hon* and to check whether it is also one of the arguments associated with the previous occurrence of *slå sönder*. If so, one must also determine whether it has the same grammatical function in this case as well. One observes that *Malin* and *Ella* are the two preceding NP's that have the same gender as *hon*, i.e. Fem. Since *Malin* is closest to the pronoun, however, it is selected as antecedent according to [15]. Examination of the syntactic functions of the first occurrence of the predicate *slå sönder* reveals that *Malin* is not its Subject; rather it is *Ella* that functions as Subject there. This is derivable from the fact that the Subject of the first occurrence of *slå sönder* must be the antecedent of *hon* in the clause *det gör hon*, since *gör* 'does' refers back to *slå sönder*. Thus, *hon* is identified as the 'new'
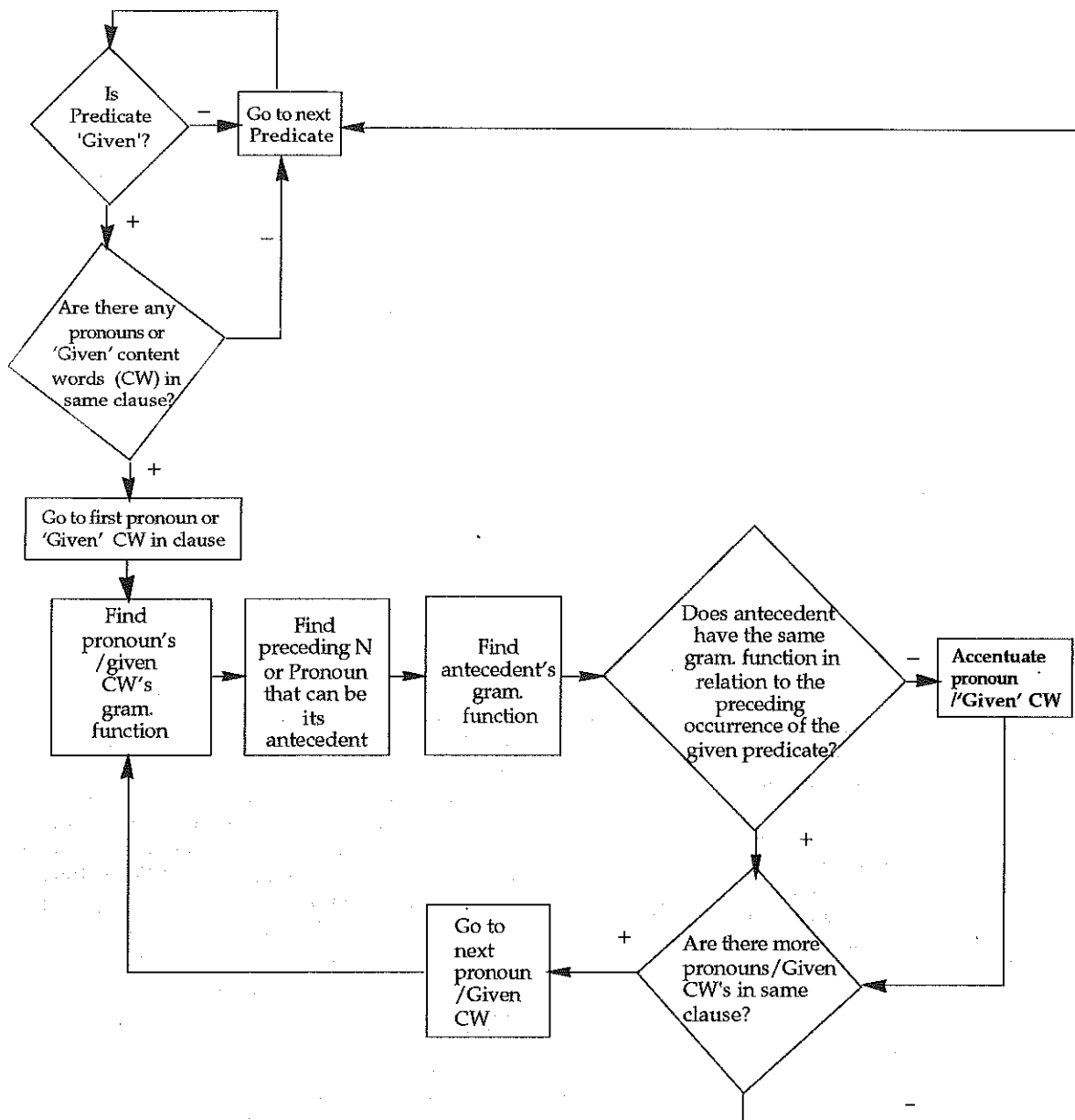
Figure 1. Flow-chart for determining conditions for accenting of personal pronouns and 'given' content words in situations of shift of grammatical function. The input is assumed to be a text which has been processed by the Referent-Tracker described in [1]. Thus the input to the present processor consists of words marked as either N(ew) or G(iven). It assumes a pronoun resolution component as in [15] as well as a procedure for determining the grammatical function of a pronoun or 'given' content word.

Subject of the second occurrence of *slå sönder* and consequently, assigned an accent.

A perhaps somewhat more complicated case of pronoun resolution is seen in the following example (taken from the Swedish translation of J. Irving's novel *Garp*, which is rich with such cases of accentuated pronouns represented with italics):

(6) Stewart Percy hade visserligen en titel, men inget verkligt jobb. Han kallades för Steeringskolans sekreterare, men ingen såg honom någonsin skriva på maskin. Faktum var att han själv hade en sekreterare,

och ingen var riktigt säker på vad *HON* kunde ha att skriva på maskin.

'Stewart Percy had to be sure a title, but no real jobb. He was called Steering School's secretary, but nobody ever saw him type. The fact was that he himself had a secretary, and nobody was really sure about what *SHE* could have to type'.

Extraction of the predicates and associated arguments in the text reveals the following linear sequence (where N = New , G = Given); underlining marks the boundary between predicate/argument groupings):

Stewart Percy	N	skriva på maskin	G
hade	N	han	G
titel	N	hade	G
jobb	N	sekreterare	G
han	G	ingen	G
kallades	N	var säker	G
sekreterare	N	hon	G
ingen	G	kunde ha	G
såg	N	skriva på maskin	G
honom	G		

Here there is a complex predicate consisting of a modal verb+infinitive *kunde ha* 'could have' and the infinitive complement *att skriva maskin* 'to type' which are given. *Hon* is thus recognized as the Subject of all these component verbs on the basis of its subject form and its position directly before the finite verbform *kunde* which is directly followed by the infinitive forms *ha* and *skriva*. The resolution of the pronoun is not as straightforward as in the previous example since there is no Proper Name whose gender agrees with *hon* (Fem.). What is required is semantic information related to the fact that *sekreterare* 'secretary' can be both feminine and masculine. The determination of the Subject of the preceding occurrence of *skriva maskin* is not straightforward either. It is construed as identical to the associated pronoun *honom* which, although in the Object form is functionally the Subject of the infinitive *att skriva på maskin* as well.

Although we have only considered the accentuation of pronouns so far, it is also the case that 'given' content words are assigned accents if they are associated with a change in grammatical function with respect to a given predicate [7] as the following example illustrates:

(7) a. Kent hit Bert [Kent>Bert]
   b. and then BERT slugged KENT

Although BERT and KENT in (7b) are 'given' entities, they are nevertheless assigned accents when they are associated with a shift in grammatical function with respect to the predicate *slugged* which is identical in sense with the previous predicate *hit*.
Both pronouns and 'given' content words can thus be processed by the same algorithm as regards accent assignment. The flow-diagram in Figure 1 summarizes the information needed in order to determine whether a pronoun or a 'given' content word should be accented or not. It is assumed that this processing would take place after referent-tracking, but before prosodic parsing.

CONCLUSION

Center shifts are sometimes correlated with specifiable formal parameters. In the case of the centered pronouns and 'given' content words, this centering often correlates with a change in grammatical function. To the extent that it is possible to resolve pronoun reference and identify grammatical functions, it will be possible to predict this centering and the associated accentuation on the basis of formal linguistic structure without recourse to more abstract intentional structure

representations. This will improve the naturalness of intonation in speech synthesis.

REFERENCES
[1] M. Horne, M. Filipsson, M. Ljungqvist, & A. Lindström. "Referent tracking in restricted texts using a lemmatized lexicon: implications for generation of prosody". *Proc. Eurospeech '93* (Berlin) Vol. 3, 2011-2014, 1993.
[2] M. Horne & M. Filipsson. "Generating prosodic structure for Swedish text-to-speech".
Proc. ICSLP 94 (Yokohama), Vol. 2, 711-714, 1994.
[3] C. Gussenhoven. "Two views of accent: a reply". *Journal of Linguistics* 21, 125-38, 1985.
[4] M. Horne. "'Contrastive prominence' and syntactic parallelism'. In *Towards a discourse-based model of English sentence stress. Working Papers* 32(Dept. of Ling., U. of Lund), 1987.
[5] M. Horne. "Accentual patterning in 'new' and 'given' subjects in English". *Working Papers* (Dept. of Ling., U. of Lund) 36, 81-97, 1990.
[6] M. Horne. "Why do speakers accent 'given' information?". *Proc. Eurospeech 91* (Genova), 1279-1282, 1991.
[7] J. Terken & J. Hirschberg. "Deaccentuation of words representing 'given' information: effects of persistence of grammatical function and surface position". *Language and Speech* 37, 125-45, 1994.
[8] C. Nakatani. "Discourse structural constraints on accent in narrative". *Proc. ESCA Workshop on Speech Synthesis*, New Paltz, N.Y., 228-231, 1994.
[9] C. Nakatani . "Accenting on pronouns and proper names in spontaneous narrative". *Proc. ESCA Workshop on Prosody*, Lund. *Working Papers* (Dept. of Ling., U. of Lund) 41, 164-167, 1993.
[10] B. Grosz, A. Joshi & S. Weinstein. "Providing a unified account of definite phrases in discourse". *Proc. 21st Annual Meeting of the Association of Computational Linguistics*, 44-50, 1983.
[11] B. Grosz, A. Joshi & S. Weinstein. "Centering: a framework for modelling the local coherence of discourse". *Computational Linguistics* 21 (in press), 1995.
[12] N. Enkvist. "Marked focus: functions and constraints. *Studies in English linguistics for Randolph Quirk*, ed. S. Greenbaum, G. Leech & J. Svartvik, 134-52. London: Longmans, 1979.
[13] G. Lakoff. "The role of deduction in grammar". *Studies in linguistic semantics*, ed. C. Fillmore & D. Langendoen, 62-70. New York: Holt, Rinehart & Winston, 1971.
[14] M. Kameyama. "Stressed and unstressed pronouns: complementary preferences". *IBM Working Papers of the Institute for Logic and Linguistics* 8 (Focus and Natural Language Processing, Vol. 3), 475-484, 1994.
[15] K. Fraurud. "Pronoun resolution in unrestricted text". *Nordic Journal of Linguistics* 11, 47-68, 1988.