Improving the Prosody in TTS Systems: Morphological and Lexical-Semantic Methods for Tracking 'New' vs. 'Given' Information

Merle Horne*, Marcus Filipsson*, Christer Johansson*, Mats Ljungqvist‡ and Anders Lindström‡
*Dept. of Linguistics and Phonetics, University of Lund, Helgonabacken 12, S-223 62 Lund, Sweden
‡INFOVOX, Box 2069, S-171 02 Solna, Sweden

ABSTRACT

The design of an algorithm for referent tracking in a restricted domain is presented. The algorithm allows one to preprocess a text and automatically tag words as either contextually 'New' or 'Given'. The procedure involves computational modelling of lexical semantic identity of sense relations as well as information on inflexional/derivational morphology and compounding. Referent identity is defined on head-word representations derived from the text input on the basis of the inflexional expansion rules contained in a lemmatized lexicon of Swedish. Information on the New/Given status of words can subsequently be used in the F_0-generating component of the text-to-speech system to trigger the assignment of focal vs non-focal word accents.

INTRODUCTION

A major goal in current research in text-to-speech has involved improving the prosody component by developing interfaces which extract contextual and syntactic information that condition pitch accent-type as well as level of accentual prominence (Hirschberg 1990, Youd and House 1991, Horne and Johansson 1991, in press, Horne, Filipsson, Ljungqvist and Lindström 1993, Monaghan 1990). In most commercial text-to-speech systems, the prosodic component does not have access to higher level syntactic and semantic information and is therefore able to generate only a very restricted number of intonation patterns. In the case of Swedish (Carlson and Granström 1973, Bruce and Granström 1989), the commonly used method is to assign a focal accent to the last content word in an intonational phrase. This focal accent pattern leads to the interpretation of all phrase-final words as 'new' information. It has not been possible to *automatically* assign phrase-final content words non-focal accents which are associated with 'given' information. In order to enable automatic assignment of the proper accent-type, it is necessary to process the input text with respect to the information status associated with the 'content' (lexical) words. We have currently been involved in developing such a linguistic preprocessor which models and tracks morphological and lexical-semantic coreferential relationships between content words (Horne, Filipsson, Ljungqvist and Lindström 1993). In what follows, we will describe how the processor works.

METHODOLOGY

In the modelling of the identity relationships and the development of the coreferent tracking algorithm, we have been currently exploiting the information contained in a computerized lexicon of Swedish (Hedelin, Jonsson and Lindblad 1987). The lexicon, which is lemmatized, contains approximately 116,000 headwords, each one listed with its part-of-speech specification, inflection code, and phonetic transcription. It also includes information on the morphological status of derivations and can handle the analysis of compound-words, either by explicit listing or by algorithmic generation.

A general feature of a lemmatized lexicon is the inherent relationships between the head-words and their inflected forms. As an example, the word *tända* 'to light' can be related to its paradigmatic forms: *tända* (inf.), *tänder* (pres.), *tände* (pret.), *tänt* (past part.), *tänd* (supine), *tändas* (inf. passive), etc. This is important in the present application since the identity relations are defined over the stems or head-words. The lexical structure has furthermore been amended with domain-specific knowledge describing semantic hierarchies (hyponymy, part/whole), synonymy relationships as well as pragmatically/situationally Given terms.

In the initial stages of the development of the algorithm, we did not have recourse to a lexicon (Horne and Johansson 1991, in press) and thus the tracking process was more limited as regards the number of different kinds of coreference relations that could be identified. Stem identification was achieved by means of a morphological truncation procedure (due to B. Brodda). This process applied each time a word was compared with each preceding word when determining its coreferential status. Morphological truncation is based on graphic information and searches for identical strings of letters in two words. If these are found and if the remaining strings in the two words are existing endings in the language, then the words are classified as coreferential. This method suffers, however, from the fact that it can only handle non-suppletive paradigms. Stems of morphs in suppletive paradigms such as *falla /föll* 'to fall/fell' cannot be found using the truncation procedure. Analysis of compounds is another problem area which was not solved using this non-lexical method.

DESIGN OF THE ALGORITHM

In the analysis of an input text, the following steps are currently involved. First, the lexicon is used to analyse the words and decompose them into morphs (see Figure 1). As mentioned earlier, the lexicon handles inflexions, derivations, and compound words. The treatment of compounds is an important feature, since the referent-tracking processes must apply not only to the compound as a whole, but to the component morphs. This process is complicated in Swedish by the fact that compounds, as in German, are written as undivided words, without hyphens or spaces between the component morphemes, e.g. *fondbörs* 'stock-exchange' consists of the morphemes *fond* and *börs*. Since the lexicon contains information on the internal structure of compounds, the tracking procedure can apply to the individual component morphemes. In Figure 2 can be seen an example of the decomposition of words into morphs.

After analysing each word in the text into its basic morph(s), the referent tracking procedure can apply to the text. Each word is then checked for its possible coreference with any previously mentioned word within an adjustable window. The window used in the examples in this paper has somewhat arbitrarily been chosen to be 60 words, but other domains could also be considered, e.g. the paragraph (Hirschberg 1990). The referent-tracking algorithm consists of four parts, as shown in Figure 1 (a-d):

The first one tracks and marks words that are situationally/pragmatically 'given', such as *börs* 'stock-market', and *krona* 'crown' in the domain we have studied (Swedish stock-market). The second stage involves identifying cases of coreference due to reiteration of root morphs (or stems), as obtained from the initial analysis using the lexicon. The third part uses 'domain-specific' synonymy relations to identify cases of coreference. Examples from the stock-market domain are: *kurs–nivå* 'rate' and *aktie–papper* 'share'. The fourth and final stage attempts to track words that are involved in hierarchical identity relations (hyponymy, part/whole relationships). In order to do this, superordinate relations have been modelled using 'is an example of' or 'is a part of' pointers to establish the relation between pairs of lemmas thus building up a forest of hierarchical, multi-branch trees. Cases of multiple inheritance, i.e. where a lemma has more than one parent (e.g. both *vardag* 'week-day' and *arbetsdag* 'work-day' are parents of *måndag* 'Monday') do not pose a problem, since the algorithm searches only among 'daughters' of the currently analyzed word, i.e. the tracking in semantic hierarchies is unidirectional, from the more general anaphor to the more specific antecedent. Output from the text processing scheme as described above is shown in Figure 3. Each word is marked as either 'new' (N) or 'given' (G).

DISCUSSION

The algorithm described in the present paper analyses the input text and annotates it with respect to 'new' and 'given' information, thereby allowing the TTS system to generate synthetic prosody of considerably improved naturalness as compared to the default behaviour, as verified by informal listening tests.The performance of the processing can, however, be improved by further analysis, such as identification of syntactic phrase boundaries (Bruce, Granström, Gustafson and House 1992) as well as relative levels of pitch prominence. The lexicon can be exploited at this stage as well in order to extract information on e.g. word-class designation. Contrastive prominence is another phenomenon which requires lexico-syntactic information. Thus, an even more attractive approach is to integrate the linguistic/contextual processing into the text-to-speech system (Lindström, Ljungqvist, and Gustafson 1993), thereby allowing it to exchange information with other knowledge sources of the TTS system, such as the syntactic parser, punctuation ambiguity resolution, treatment of numbers and abbreviations, etc.

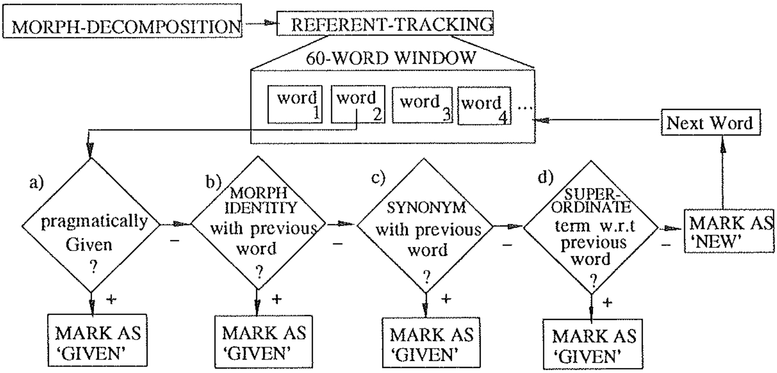

Figure 1. *Flow-diagram illustrating the different stages of the analysis in the lexical processor: MORPH-DECOMPOSITION first analyses each word into its content morphs. In the second stage, the REFERENT-TRACKING algorithm searches through a 60-word window for coreferents.*

:Stockholms	'Stockholm's'	:slutade	'closed'
Stockholm	'Stockholm'	sluta	'close (inf.)'
:fondbörs	'stock-exchange'	:på	'on'
fond	'stock'	på	'on'
börs	'exchange'	:torsdagen	'Thursday+def.art'
:generalindex	'general index'	torsdag	'Thursday'
general	'general'	:på	'at'
index	'index'	på	'at'
		:858.8	'858.8'
		858.8*	'858.8'

Figure 2. *Example of decomposition of words into morphs using the lemmatized lexicon. Analysed words are preceded by a colon (:). Under each of these words is/are the component root-morph(s). Notice that for verb-forms, the infinitive form is taken to be the root. An asterisk following a form indicates that there is no corresponding lexical entry.*

```
 1N   Stockholms   12N    marginella   23N    som              34Gw och
 2Gg  fondbörs     13N    0.02         24N    oregelbunden     35G29 läkemedel
 3N   generalindex 14N    procent      25Gs 10 Kursstegringar  36Gw att
 4N   slutade      15N    jämfört      26N    i                37N  bli
 5N   på           16Gi 11 med         27N    AGA              38N  bästa
 6N   torsdagen    17N    onsdagens    28Gw   och              39Gi 31bransch
 7Gi 5 på          18Gi 3 slutindex    29N    Astra
 8N   858.8        19N    Kursutveckl.30N     fick
 9N   en           20N    över         31Gi 3 branschindex
10N   uppgång      21Gh 6 dagen        32N    för
11N   med          22N    betecknades  33Gh 27 kemi
```

Figure 3. *An example of the output from the referent-tracking algorithm. Words are marked either as N for 'new' or G for 'given'. If a word is marked as G, the output shows in addition the kind of givenness that is present: Gg stands for pragmatically given, Gi, for 'given' due to morphological identity. The number directly preceding the word refers to the number of the word to which the word is construed as coreferent. Gh stands for 'given' due to a hierarchical relationship, e.g. word 21, dagen is marked as Gh since it refers back to word 6, torsdagen. Gs refers to coreference due to synonymy, e.g. the component stegring 'rise' in the compound word 25 kursstegringar 'rate-increases' is a synonym to word 10, uppgång, thus triggering a non-focal accent on the compound. Gw stands for grammatical words that are not eligible to receive focal accents.*

ACKNOWLEDGEMENTS

This research has been supported by a grant from the Swedish HSFR/NUTEK Language Technology Programme.

REFERENCES

G. Bruce and B. Granström (1989), "Modelling Swedish intonation in a text-to-speech system", *STL-QPSR*, Vol. 1, pp. 31-36.

G. Bruce, B. Granström, K. Gustafson and D. House (1992), "Aspects of prosodic phrasing in Swedish", *Proc. ICSLP, Banff, Canada,* Vol. 1, pp. 109-112.

R. Carlson and B. Granström (1973), "Word accent, emphatic stress, and syntax in a text-to-speech system, *STL-QPSR*, Vol. 2-3, pp. 31-36.

P. Hedelin, A. Jonsson and P. Lindblad (1987), *Svenskt uttalslexikon: 3 ed. Tech. Report, Chalmer's Univ. of Technology.*

J. Hirschberg (1990), "Using discourse context to guide pitch accent decisions in synthetic speech", *Proc. ESCA Workshop on Speech Synthesis, Autrans, France,* pp. 181-184, .

M. Horne and C. Johansson (1991), "Lexical structure and accenting in English and Swedish restricted texts", *Working Papers (Dept. Ling., U. of Lund),* Vol. 38, pp. 97-114.

M. Horne and C. Johansson (in press), "Computational tracking of 'New' vs 'Given' information: implications for synthesis of intonation", *Proc. Nordic Prosody VI, K.T.H., Stockholm, August 12-14, 1992.*

M. Horne, M. Filipsson, M. Ljungqvist and A. Lindström (1993), "Referent tracking in restricted texts using a lemmatized lexicon: implications for generation of intonation", *Proc. Eurospeech '93, Berlin, 21-23 September, 1993.*

A. Lindström, M. Ljungqvist and K. Gustafson (1993), "A modular architecture supporting multiple hypothesis for conversion of text to phonetic and linguistic entities". *Proc. Eurospeech '93, Berlin, 21-23 September, 1993.*

A. Monaghan (1990), "Treating anaphora in the CSTR text-to-speech system". *Proc. ESCA Workshop on Speech Synthesis, Autrans, France,* pp. 113-116.

N. Youd and J. House (1991), "Generating intonation in a voice dialogue system", *Proc. Eurospeech 91, Genua, Italy,* pp. 1287-1290.