PROSODIC CORRELATES OF INFORMATION STRUCTURE IN SWEDISH HUMAN-HUMAN DIALOGUES

Merle Horne, Petra Hansson, Gösta Bruce and Johan Frid
Dept. of Linguistics and Phonetics, Lund University, Helgonabacken 12, S-223 62 Lund, Sweden
First Name.Last Name@ling.lu.se, http://www.ling.lu.se

ABSTRACT

Investigation of travel agent (=System) Task-related utterances reveals intonational contours where focal accents occur on information given previously in the dialogue by the client (=User). The accentuation can be related to the interactive nature of dialogue where the information holder picks up on a Task-related topic introduced by the client and comments on it in an engaged manner.

Keywords: prosody, accentuation, information structure, dialogue, speech synthesis, Given-New

1. INTRODUCTION

Within the Swedish Dialogue Systems project (http://www.ida.liu.se/~nlplab/sds/) involving five different research groups within Sweden, the principle aim is to pool expertise represented at these different locations in order to develop a common environment which will allow for the creation of generic dialogue systems. The focus is on the higher system levels: dialogue management, robust utterance interpretation, including use of prosodic information, and effective utterance planning.

The subproject in Lund (Prosody, Grammar and Discourse Structure) has as its goal to investigate how prosodic information interacts with lexical and syntactic information in order to:

1) mark boundaries between different kinds of dialogue moves and
2) to express the contrast between New information and different kinds of Given information.

As regards the first area, we have been particularly interested in getting a better understanding of how discourse markers, together with their prosodic patterns and local lexical context can signal the beginning of a new topic as well as the beginning of different kinds of dialogue moves [4], [7].

The current study is related to the second goal of the Lund subproject, where we are interested in getting a better understanding of how information structure in dialogue is expressed in terms of different kinds of linguistic form (prosodic, lexical, syntactic).

2. INFORMATION STRUCTURE AND PROSODY

In earlier work on prosody in discourse, it has been shown how prosody (in particular, accentuation) interacts with lexical semantic relations in the expression of coreferents /cospecifiers, i.e. Given information [6]. Although it now seems clear that Given information is deaccented in phrase-final position, it is also clear that it is not always deaccentuated in phrase-initial position. Various kinds of information have been brought to bear in an attempt to explain this phenomenon, e.g. rhythmical constraints which require an early prominence in a phrase [5] or discourse constraints related to marking a change in a referent’s thematic status. This can be expressed in terms of a shift in grammatical function as well as accentuation [9]. Subject position, for example, is known to be a position for encoding thematic referents and attributes [10]. In terms of the notion of ‘grounding’ (the process of adding to common ground) [3], the initial subject can constitute the ‘link’ to Given information [11]. In the context of spoken dialogue, this grounding occurs in an interactive way when one dialogue participant links back to information introduced first (Given) by another dialogue participant and elaborates on it in some way.

3. ACCENTUATION AND THE GIVEN-NEW DISTINCTION IN SWEDISH

Swedish is a pitch accent language with two lexical word accents, Accent 1 and Accent 2. According to Bruce [2], these two accents are distinguished from each other as to their timing relative to the segmental string. Accent 1 is represented in the standard (Stockholm) dialect as HL* whereas Accent 2 is represented as H*L, where the starred tone is critically aligned with the beginning of the vowel in the stressed syllable (see Figure 1). In addition to its word accent, a word is also assigned a prominent phrasal tonal gesture consisting of a high tone (H-) followed by a phrasal tone (L%) when the word is made optimally prominent (focussed). Thus a New Accent 1 word will have the ‘underlying’ representation HL*H*L% and an Accent 2 word the representation H*LH*L%. For focal Accent 1 words which are often monosyllabic, it is very often only the focal tone and phrase accent (H-L%) which are realized due to the lack of segmental content. It is still
distinguished from Accent 2, however, since Accent 2 words always have at least two syllables and thus allow realization of the entire tonal contour.

![Schematic representation of F0 contours on focussed Accent and Accent 2 words following Bruce](image)

**Figure 1. Schematic representation of F0 contours on focussed Accent and Accent 2 words following Bruce [2].**

### 3.1 Accentuation of Given Information in Swedish

As mentioned in section 2, it has been observed that Given information is sometimes prominently (focally) accented in phrase-initial subject position. What has also been observed in regard to these very prominent accents on Given information is that the timing of the focal $H^*$ with respect to the segmental string appears to be different than when the associated word is New. Briefly stated, it has been observed that the focal $H^*$ tone seems to be later in Given contexts in Accent 2 words and earlier in Accent 1 words.

As far as the design of man-machine systems goes, therefore, it is of interest to know in what kind of dialogue move(s) accentuation of Given information occurs as well as how it is realized in order to a) be able to automatically recognize it in user speech and b) be able to generate it in system utterances in contexts where it is appropriate.

### 3. METHOD AND DATA

In order to get a better understanding of the context(s) where Given information is accented in spontaneous dialogue, we have begun an investigation of human-human task-oriented dialogues from the Travel domain. The data investigated consist of spontaneous, ecologically valid dialogues recorded at travel-bureaus in Lund, Sweden. As mentioned above, of particular interest is the accentuation of Given information in travel agent moves (corresponding to machine moves in a man-machine system). We compare travel agent (=System) utterances with initial (subject position) focally accented content words that are previously mentioned (Given) by the client (=User) with comparable utterances where the initial content word (both Given and New) has not been previously mentioned by the client.

### 4. TASK-ORIENTED DIALOGUE MOVE CODING

In order to interpret the intonational patterning of dialogues in a meaningful way, it is necessary to be able to relate intonation curves to dialogue move structure. In the Linköping dialogue management system being developed in the project [8], dialogue moves can be classified as Task-related, Communication Management-related or Other Contribution [1], where Task-related referents and attributes are those that are specific to a particular topic within the travel-bureau domain, e.g. Booking a Trip, Payment of a Ticket, Information Seeking, etc. Referents (‘Objects’) and attributes (‘Properties’) related to these topics can then be thought of as being organized into domain-specific hierarchical structures which are accessed and filled with specific information during the dialogue. In Figure 2 a rough sketch of how some of the dialogue topics in the Travel domain can be represented is given. During the course of a task-oriented dialogue, information on specific topics relevant to the solution of the task is retrieved. On the basis of this information, the system (travel agent) can solve the task. The labelling of the dialogue moves shown in Figure 2 reflects the distinction between Task and non-Task related information in the dialogue. The labels, taken from Ahrenberg’s coding system [1] are to be interpreted as follows:

**Task Management Functions:**

- **TO** = Task opening
- **Q** = Question
- **A** = Answer
- **E** = Elaboration
- **TS** = Task Solution

**Communication Management Functions:**

- **RI** = Repair Initiative
- **RA** = Repair

### 5. ACCENTUATION OF GIVEN INFORMATION IN DIALOGUES

A study of travel agent utterances in the project database of spontaneous human-human dialogues reveals a structured distribution of accented Given information. The accentuation would seem to be restricted to specifically Task-related referents/attributes, i.e. information which, although Given in the discourse, is
relevant to the task at hand (see section 4). Further, the Given information is moreover, information that the client (user) has previously introduced into the dialogue. Moreover, although the Given information is prominently accented, the accentual pattern would appear not to be the same as that occurring on New (focal) information (see Figure 1). Rather, the focal High (H°) after the word accent is moved to the right (see F0 contour on billigaste, Figure 3). In other words, the timing of the H° is ‘delayed’. In terms of prosodic phrasing/grouping, one can interpret this movement of the H° to a position outside the word as a way of making the accentually prominent Given word a part of the same prosodic phrase as the information that follows. New focally prominent words, on the contrary, are characterized by having the focal accent H° within their scope (see F0 contour on France, Figure 3). In contrast to non-Task oriented information in initial position, the accents on Task-related Given information are very conspicuous by their wide F0 range (see Figure 3).

6. PROSODY AND DIALOGUE INTERACTION

The wider range of the accent on initial Task-related referents and attributes can be interpreted as a specific characteristic of dialogue interaction. The travel agent takes up a topic referent first mentioned by the client and comments on it. The relative wide range of the accent can be thought of as reflecting the degree of engagement that the travel agent has in his interaction with the client.

In order to test this hypothesis, we are currently designing a perception test where we will use synthesized speech to allow manipulation of F0 range and F0-peak placement. This will allow us to see if listeners in fact prefer System utterances characterized by F0-contours of the kind described here (see Figure 3). The results are of interest for the design of the utterance generation module in man-machine spoken dialogue systems.
Figure 3. Illustration of the type of F0 contour characteristic of travel agent (system) Task-related utterances where Given information is prominently accented in initial (Subject) position. The utterance is the Swedish equivalent to the English The cheapest we have is Air France presented in Figure 1. Notice the late timing of the focal H~ which occurs in the word vi `we’ following cheapest (billigaste, an Accent 2 word). The last word in the utterance, France, has a focal Accent 1 contour. Three label tiers are shown. From top to bottom they are: 1) tonal tier with word, focal and phrase-accents represented, 2) functional tier with complex labels consisting of either G(iven) or N(ew), (task)-related status and I(nitial) or F(inal) linear position, 3) word transcription with prominence marking.

Acknowledgements

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

7. REFERENCES