PROSODIC CORRELATES OF DISCOURSE MARKERS IN DIALOGUE

Petra Hansson
Department of Linguistics and Phonetics, Lund University, Sweden
E-mail: petra.hansson@ling.lu.se

ABSTRACT

This study reports on the Swedish discourse markers *men* ‘but’/‘and’ and *så* ‘so’, their functions and prosodic correlates in dialogue. The purpose of the investigation is to get a better understanding of how discourse markers together with their prosodic characteristics signal the beginning of a new topic, a return to a previous topic, and different kinds of dialogue moves. The possibility of using prosody to automatically detect discourse markers is discussed. The results are compared with results from an earlier study on Swedish discourse markers in monologue.

1. INTRODUCTION

Schiffrin [1] defines discourse markers as sequentially dependent elements (words or phrases) which bracket units of talk, e.g. sentences or speech acts / dialogue moves. Discourse markers are used extensively in spontaneous speech and as explicit indicators of the structure of discourse by marking the beginning of a new topic, a return to a previous topic, etc [2]. Speakers also use discourse markers to signal how the upcoming utterance relates to prior discourse. They help the listener to develop an expectation of the discourse intent of the upcoming utterance [3].

This study reports on the Swedish discourse markers *men* ‘but’/‘and’ and *så* ‘so’. The purpose of the study is to get a better understanding of how discourse markers together with their prosodic characteristics signal the beginning of a new topic, the return from a digression, and different kinds of dialogue moves. The results are compared with results from an earlier study on Swedish discourse markers in monologue [4].

1.1. Swedish Dialogue Systems

The Swedish Dialogue Systems research project (http://www.ida.liu.se/~nlplab/sds/) involves five different research groups within Sweden. The main goal of the project is to develop a common environment which will allow for the creation of generic dialogue systems. The subproject in Lund, *Prosody, Grammar and Discourse Structure*, (http://www.ling.lu.se/projects/SDS/) 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 discourse units in dialogue (topics and dialogue moves) and
2) to express the contrast between ‘new’ information and different kinds of ‘given’ information.

The current study is related to the first of these goals. It constitutes a follow-up on an investigation of discourse markers undertaken during the pilot project [4], which had as its goal to investigate the possibility of using prosody to detect discourse markers in monologue.

Words that are used as discourse markers often have a sentential function as well. Sententially, *men* may be used as a conjunction linking together two clauses, and *så* may function as a clause initial place-holder (in main clauses preceded by adverbial subordinate clauses) [5]. If we can use prosody to distinguish between discourse and sentential usages, discourse markers such as *men* and *så* can be used in speech processing algorithms e.g. to facilitate the detection of discourse boundaries.

1.2. Discourse markers in spoken dialogue systems

Discourse structure identification is an important process for speech understanding systems. Information about the makeup of discourse helps language processing e.g. with the problem of reference resolution. From previous studies we know that prosody signals different aspects of discourse structure, such as topic structure, turn-taking, and type of dialogue move. The information provided by the prosodic cues can be used e.g. to segment speech [6, 7] and to identify different dialogue moves [8, 9]. Nevertheless, textual cues also constitute an important source of information which should be taken into account together with prosody. When discourse markers are present, they provide reliable cues to discourse structure [7].

Figure 1 shows a preliminary module specification of *Swedish Dialogue Systems*. Prosodic information is thought to be extracted and analyzed in the ‘Speech understanding’ module which has access to the ‘Prosodic model’. Prosodic information is as yet not included in the grammar used by the parser in the ‘Interpretation’ module. However, as the output from the ‘Interpretation’ module consists of e.g. information on possible dialogue moves and information on relevant concepts such as ‘objects’ (referents) and ‘properties’ (attributes), prosody is important to include in the interpretation grammar.

An appropriate generation of discourse markers may be useful for user inference of system intention. During the course of a task-oriented dialogue, e.g. in the travel domain, information on specific topics relevant to the solution of the task, such as type of trip, departure and return dates and means, is gathered. This information enables the system to solve the task. However, when the user does not supply the system with sufficient information on a particular topic, the system has to know how to return to it and request a clarification. In Swedish human-human dialogues, an initiative of this kind would typically be introduced by the discourse marker *men*.
Inspiration for the study on the discourse marker men in monologues [4] came from a study by Litman and Hirschberg [2] on the English marker now. Now is known as a deictic adverbial in sentence grammar, but in spoken discourse it also functions to signal the return to a previous topic. Hirschberg and Litman showed that discourse now differed from sentential (adverbial) now in that it, e.g., often appears at the beginning of a phrase or constitutes a prosodic phrase on its own, more often than sentential now is deaccentuated and often is preceded by another discourse marker (well now, okay now, right now).

In [4], we examined the prosodic and lexical correlates of the Swedish marker men in monologues. Only the results from the study of the tokens whose classification all of the authors agreed upon (using textual information) are presented here. These cases involve “strong” boundary marking where “boundary strength is computed as the proportion of subjects agreeing on a given break” [10: 515]. It was argued that it is these clearly marked cases that should be important for language processing since they are the cases the speaker probably intends the listener to pay particular attention to.

The results showed that the prosodic characteristics of men had a good disambiguating power. Mean differences in F0-reset and absolute word duration were significant. A difference in prosodic phrasing was also observed. The discourse tokens of men were further associated with a local lexical correlate, i.e. discourse men were in over 60% of the cases followed by another discourse marker. The question arose whether discourse markers in dialogue were associated with the same prosodic characteristics.

3. CURRENT STUDY

3.1. Data

Four spontaneous, ecologically valid dialogues were recorded at two different travel agencies in Lund, Sweden. The dialogues are between 2 and 15 minutes long, yielding about 23 minutes of speech in total.

3.2. Labeling

All together the dialogues included 47 tokens of men and 67 tokens of så. Using both textual and auditory information, the tokens of men and så with a discourse function were distinguished from those with a sentential function. The discourse men and så were labeled D and the sentential men and så were labeled S.

3.3. Acoustic analyses

The recorded speech was segmented and orthographically transcribed in label tiers (synchronized with the speech wave and F0 contour) using the ESPS/waves+™ analysis program. Subsequently, the prosodic correlates of both discourse and sentential tokens of men and så were investigated and compared with each other.
Pauses. Measurements were made of the duration of pauses preceding turn-internal men and så. Both pauses located immediately before men and så and pauses located before a discourse marker or a hesitation sound occurring between men and så and the silent pause were measured.

One would expect relatively longer pauses before discourse men and så than sentential men and så since the discourse tokens mark the beginning of a new prosodic utterance while sentential men and så often occur at the start of a new prosodic phrase. Prosodic utterances (‘speech paragraphs’ or ‘discourse units’) are most often preceded by longer pauses than prosodic phrases (often corresponding to a clause in written language) [11].

The general distribution of both turn-internal and turn-initial pauses before and after discourse and sentential uses of men and så was also investigated.

Word duration. Measurements were made of the absolute duration of men and så. Conjunctions (men) and place-holders (så) are non-prominent word classes. However, when functioning as discourse markers, they are expected to be drawn out as a result of the speaker’s need for extra speech planning time when, e.g., introducing a new topic.

 estão reset. Measurements of estão were made at the end of the phrases preceding men and så and in the vowels of men and så. No measurements were made in glottalized parts of the speech. The expectation would be to find a relatively larger estava reset on D-men and D-så than on S-men and S-så since the degree of topic-continuity between what precedes and follows discourse men and så (i.e. different discourse units) is less than between what precedes and follows sentential men and så (i.e. clauses pertaining to the same sentence) [12].

4. FUNCTIONS OF DISCOURSE MARKERS IN DIALOGUE

4.1. Men ‘but’/‘and’

Our prior work on the Swedish discourse marker men in monologues [4] suggests that it has two main functions: 1) as a marker of the beginning of a new topic, and 2) as a marker of the return to a previous topic, as in (1) and (2).

(1) ja han är förtvivlad verkar det som åh sedan blir han mer eller mindre utkörd från hotellet men åh så åh väntar han tills det blir kväll
   ‘well he’s devastated it seems like uh then he’s more or less thrown out of the hotel and uh then uh he waits until it’s night’

(2) jag vet inte om han stannar kvar på hotellet eller om han lyckas ta sig ut därför eller hur det nu var men i alla fall så går han kommer tillbaka
   ‘I don’t know if he stays at the hotel or if he manages to get away from there or what happened but anyway he leaves returns’

In the examined dialogues, men is frequently used when introducing a new topic (e.g. in Ready moves [13]), as in (3).

(3) men då så då är det då måste vi bara ta och åh fixa fram en biljett här åt dig
   ‘okay then then it’s then we’ll just have to uh get you a ticket here’

It is also found at the beginning of turns (mainly in Explain and Query-YN moves [13]) which express a contrast or constitute an objection to something which has been said earlier on in the dialogue, as in (4). This discourse meaning of men is quite similar to the meaning of the sentential men ‘but’.

(4) A: så det man kan göra om man hittar platser alltså för det kan vara rätt så fullobokat i och med att det är över jul nyår ja
   B: ja men om man inte vet priset så känns det ju
   A: ‘so what you can do if we find any seats that is because it can be pretty booked up since it’s around the holidays right’
   B: ‘yes but if you don’t know the price then it feels’

4.2. Så ‘so’

The marker så has several discourse functions. It can be found in logical sequences, at the beginning of an utterance which constitutes a conclusion, or a request for confirmation of a conclusion, e.g. in a Check move, as in (5).

(5) så den här ska jag bekräfta
   ‘so I’ll have to confirm this’

Så is also found in temporal sequences (Explain and Instruct moves), where it is interchangeable with sedan ‘then’, as in (6).

(6) och så skulle jag be dig fakturera på en universitets-institution
   ‘and then I’d like to ask you to bill a department at the university’

Så has an additional function which is closely related to its function as a discourse marker. It serves as a topic place-holder (TPH) after topicalized adverbs and discourse markers, as in (7).

(7) och sedan så är man då i Köpenhamn på morgonen den femte
   ‘and then TPH you’re in Copenhagen in the morning on the fifth’

One can expect that the above mentioned discourse functions of men and så would be associated with prosodic cues which help listeners interpret them as markers of the discourse structure, and which could be used to detect discourse markers automatically.

5. RESULTS

The results of the prosodic analysis show that different prosodic correlates characterize the two functions of men and så. Furthermore, the correlates of D-men are similar to those of D-så and the correlates of S-men similar to those of S-så.
However, there are some differences as regards the prosodic characteristics of the D-men tokens in the dialogues and the D-men tokens in the monologue data studied in [4].

5.1. Prosodic cues

Pauses. In 73% of the cases (68 of 93 tokens), the discourse men and sà are preceded by a silent pause. Seven tokens of men and sà, all functioning as discourse markers, are both preceded and followed by pauses and constitute phrases on their own. Moreover, five discourse men and two discourse sà are followed by filled pauses (hesitation sounds). Of the 21 tokens of sentential men and sà, only 3 (14%) are associated with pauses.

Pauses can be seen as either signals to improve comprehension or traces of production difficulties [14]. Either way, one would expect that the pauses preceding discourse men and sà would be relatively longer than pauses preceding sentential men and sà.

Pauses intentionally introduced to signal the presence of a boundary are expected to vary in duration depending on the type of boundary they mark. When the purpose of the pause is to signal a topic shift, as in the case of pauses preceding discourse men and sà, one expects the pause to be relatively longer than a pause marking e.g. a prosodic phrase boundary [11]. A pause can also be a trace of speech planning. One can assume that a speaker encounters some production difficulties when introducing a new topic which could cause the time needed to initiate speech to increase. Since discourse markers, but not conjunctions and clause-initial place-holders, often are found in association with topic changes, one expects to find relatively longer pauses before discourse than sentential men and sà.

However, because of the few occurrences of pauses that precede sentential men and sà in the examined dialogues, it is not possible to determine whether there in fact is a difference in the length of the preceding pauses. The measurements of the pauses’ duration (only turn-internal pauses) show that S-men are preceded by pauses which are on the average of 560 ms long (n=2), D-men of pauses 540 ms long (n=32, SD=580 ms), S-sà of pauses 480 ms long (n=1), and D-sà of pauses 970 ms long (n=26, SD=1130 ms). Nevertheless, discourse men and sà are more often than sentential men and sà associated with pauses both preceding and following them, as shown in Table 1.

<table>
<thead>
<tr>
<th>Pauses</th>
<th>Before</th>
<th>After</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentential men</td>
<td>40% (2)</td>
<td>0% (0)</td>
</tr>
<tr>
<td>Discourse men</td>
<td>90% (38)</td>
<td>17% (7)</td>
</tr>
<tr>
<td>Sentential sà</td>
<td>6% (1)</td>
<td>0% (0)</td>
</tr>
<tr>
<td>Discourse sà</td>
<td>59% (30)</td>
<td>8% (4)</td>
</tr>
</tbody>
</table>

Table 1. Distribution of pauses preceding and following men and sà. The number of tokens is given in brackets.

In the monologue data studied in [4], a tendency towards longer pauses before D-men than S-men was observed, as shown in Table 2. The difference in average duration was not statistically significant (p>0.05), however. As in the dialogues, pauses preceding D-men tokens (66%) were more common than pauses preceding S-men tokens (38%).

<table>
<thead>
<tr>
<th>Pauses</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentential men</td>
<td>460 (15)</td>
<td>350</td>
<td>40-1080</td>
</tr>
<tr>
<td>Discourse men</td>
<td>750 (27)</td>
<td>670</td>
<td>30-2320</td>
</tr>
</tbody>
</table>

Table 2. Duration of pauses preceding discourse and sentential men in the monologue data. The number of tokens used to calculate the duration statistics is given in brackets.

Word duration. Measurements of the absolute duration of men and sà show no statistically significant difference in average duration between the D and the S labeled tokens. Both discourse and sentential men and sà are relatively short in the examined dialogue data.

<table>
<thead>
<tr>
<th>Word duration (ms)</th>
<th>Mean</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sentential men (5)</td>
<td>120</td>
<td>70</td>
<td>50-240</td>
</tr>
<tr>
<td>Discourse men (42)</td>
<td>150</td>
<td>80</td>
<td>70-370</td>
</tr>
<tr>
<td>Sentential sà (16)</td>
<td>120</td>
<td>30</td>
<td>50-170</td>
</tr>
<tr>
<td>Discourse sà (51)</td>
<td>150</td>
<td>130</td>
<td>50-700</td>
</tr>
</tbody>
</table>

Table 3. Duration of discourse and sentential men and sà. The number of tokens used to calculate the duration statistics is given in brackets.

However, as seen in Table 3, there are a number of occurrences of discourse men and sà with a relatively long duration. 14% of the D-men and D-sà tokens have a longer duration than the longest tokens of S-men and S-sà, which suggests that a long duration in some cases is a prosodic characteristic of discourse markers.

In the monologue data [4], duration differences between D-men and S-men were statistically significant (p=0.0020, t=3.2, df=78). The average duration of D-men tokens was 310 ms (n=41, SD=240) and S-men tokens 170 ms (n=39, SD=130). Note, however, that the average (absolute) duration of S-men tokens in the monologue data is about the same as the average duration of D-men tokens in the dialogue data. This may be an effect of a speech rate difference between the two speaking styles or, since many of the discourse markers in the dialogues occur turn-initially, a result of the speaker planning his or her next contribution during the other speaker’s turn in dialogues. Either way, it represents a problem if word duration is to be used for automatic detection of discourse markers.

F₀ reset. Measurements of F₀ show that a larger mean positive F₀ reset is associated with discourse men and sà than sentential men and sà. D-men tokens are associated with a mean reset of 3.73 ST (SD=3.08) and S-men with a reset of 1.75 ST (SD=1.03). D-sà is associated with a mean reset of 5.13 ST (SD=4.16) and S-sà with a reset of 1.99 ST (SD=2.57). The difference in mean reset is statistically significant (p=0.0218, t=2.370, df=49) in the case of sà, but not in the case of men (p>0.05).
The difference in positive $F_0$ reset was significant ($p=0.0448$, $t=2.093, df=30$) also in the monologue data [4]. The mean $F_0$ reset in association with D-men was 5.7 ST (SD=5.6) and S-men 2.2 ST (SD=3.6).

Low $F_0$ values and creak (glottalization) was more often observed in association with discourse than sentential men and så (in 15 of 16 cases). The use of glottalization to mark prosodic boundaries is well known. Dilley, Shattuck-Hufnagel, and Ostendorf [15], for example, have shown that speakers are more likely to glottalize word-initial vowels when those vowels occur at the beginning of a new intonational phrase. The results of the present study suggest that speakers of Swedish also use glottalization as a marker of discourse structure. Moreover, vowels associated with boundaries of higher-ranked prosodic constituents, such as the prosodic utterance boundary (where discourse markers are found), seem to be more frequently glottalized than vowels associated with boundaries of lower-ranked prosodic constituents.

The same tendency was found in the monologue data [4]. 17 of the 41 D-men tokens (42%) were glottalized as opposed to only 9 of the 39 S-men tokens (23%).

### 5.2 Additional cues

In the examined dialogues, 33% of the discourse men and så are followed by another discourse marker or a hesitation sound (31 of 93 tokens). As predicted, sentential men and så are never followed by discourse markers. A following possible discourse marker (i.e. a word which may be used as a discourse marker) thus constitutes a reliable and very useful cue for identification of discourse markers.

Following are some examples of common combinations in which men and så occurred in the material: jo men alltså, men då (så), och så, och sedan (så), så då (så) (corresponding approximately to English 'yes but', 'but then', 'and then': 'so then'). Although only following discourse markers have been studied here, preceding ones can also be used for disambiguation of potential discourse markers.

A large percentage of the D-men tokens in the monologue data (63%) were also followed by another discourse marker. However, as a result of using a larger window in the automatic search for possible following discourse markers (a window of 5 instead of 3 words following the men token), some of the S labeled tokens appeared to be followed by discourse markers as well. It turned out, however, that these ‘potential’ discourse markers are not in fact functioning as such.

### 6. CONCLUDING REMARKS

Summarizing the results of this investigation, it appears that there are several prosodic and lexical cues which help a listener to disambiguate men and så and decide whether a discourse or a sentential use is meant. These cues may prove useful for automatic detection of discourse markers. As far as associated prosodic correlates are concerned, discourse men and så were characterized by a larger $F_0$ reset than sentential men and så and in some cases also with a longer duration than sentential men and så. As regards lexical cues, it was seen that a third of all discourse men and så were followed by another discourse marker. Preceding discourse markers were also common and could also be used for disambiguation of potential discourse markers.

Due to the short turns in the investigated dialogues, most of the tokens of discourse men and så occurred in turn-initial position (which is reflected in the large number of discourse men and så that are preceded by pauses). The turn-initial position therefore provides an additional, reliable cue for identification of men and så’s discourse functions in dialogue. However, not all discourse markers occur in turn-initial position, nor are all discourse markers followed by another discourse marker or have all of the prosodic characteristics mentioned above, see Figure 2.

Figure 2 shows an example of a token of men which has been labeled D, since it marks the return from a digression. It is associated with a relatively large $F_0$ reset, and it is preceded by a silent pause. Furthermore, it is both preceded and followed by words which also function as discourse markers. However, the short duration of men is not an expected characteristic of a discourse men, nor is the short duration of the preceding pause. The preceding pause’s duration and the word duration rather seem to indicate that a sentential usage was meant, although this was not the case.

Not all cues are necessary or equally important for the interpretation of men and så. Therefore, a number of features need to be taken into consideration when disambiguating potential discourse markers.

Using the monologue data, a prosody-based classifier was built with a neural network [4]. It used three input nodes for the parameters ‘preceding pause duration’ (ppd), ‘absolute word duration’ (dur) and ‘$F_0$ reset’ ($F_0$). The net had a hidden layer with two nodes, and an output layer with one node for the classification as D-men or S-men. The neural network processing and simulation was performed with the SNNS [16] package from IPVR in Stuttgart and the training was performed using the Resilient back propagation (Rprop) scheme. The network was trained in two different sessions which ran independent of each other. The results, an average of the two sessions, showed a 90% rate of correct classifications [4].

However, the comparison of the prosodic characteristics of the discourse marker men in monologue and in dialogue, e.g. as regards word duration, shows that further input nodes need to be added to the classifier in order to successfully detect
discourse markers in dialogue. The presence of a preceding pause, a preceding or following discourse marker and the position of men or sá within the turn (initial / non-initial) are parameters which, in combination with the prosodic parameters, will prove useful.

7. ACKNOWLEDGMENTS

The research reported on here has been supported by a grant from the Swedish HSFR/NUTEK Language Technology Program.

8. REFERENCES


