Silence and Discourse Context in Read Speech and Dialogues in Swedish

Sofia Gustafson-Capkova & Beàta Megyesi
Department of Linguistics, Stockholm University, Sweden & Centre for Speech Technology, KTH, Stockholm, Sweden

Abstract
We investigate the correlation between silent intervals (pauses) and discourse boundaries in the notion of theme shift. A considerable attention is given to a) the syntactic and discourse context in which pauses appear b) the characteristics of the discourse structure in terms of pauses. A case study of three speaking styles with four speakers in each style is carried out.

Definitions
Pause is defined as a silent interval equal to or longer than 100 ms.
1. Silent: silence in the speech signal
2. Complex: may include breathing and swallowing (natural physical phenomena) but not humming or other types of meaningful feedback (back channeling).
Theme is defined as a chunk with one underlying intention, that is a sequence of utterances aiming to communicate the same intention
Theme shift (TS) is the position in the discourse where a new theme is introduced and marks a discourse boundary.

Speaking styles
• spontaneous speech: elicited dialogues
• professional news reading
• non-professional news reading

Data
Swedish radio news - 920 words each (Sveriges Radio)
* read by 4 professional announcers (2 male, 2 female)
* read by 4 non-professional readers (2 male, 2 female)
Swedish map task dialogue - 920 words (Helgason)
* two dialogues with two participants (1 male, 3 female)

Method
• Pauses: Automatic detection of pauses in audio data.
• Theme shift: 5 subjects annotation of discourse structure with regard to theme shift in transcription of the data.
• Syntactic annotation: The transliterated texts were annotated automatically with part-of-speech including morphology, and shallow syntactic analysis (NP, AP, ADVP, PP, NPMASS, etc)

RESULTS

Duration and frequency of pauses

<table>
<thead>
<tr>
<th>Speaking style</th>
<th>Pause duration</th>
<th>Word/Pause Ratio</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional</td>
<td>271 ms</td>
<td>77 (920/12)</td>
</tr>
<tr>
<td>Non-professional</td>
<td>561 ms</td>
<td>8.4 (920/110)</td>
</tr>
<tr>
<td>Dialogue</td>
<td>538 ms</td>
<td>5.5 (920/167)</td>
</tr>
</tbody>
</table>

Recall and Precision for pauses and theme shift

Recall
the percentage of acoustic pauses occurring at theme shift

Precision
the percentage of TS that corresponds to silence

Professional reading:
Every pause is located at TS (100% recall)
Few TS is marked as a pause (14% precision)

Non-professional reading:
Most of the pauses at TS (71% recall)
Most of the TS are marked as a pause (92% precision)

Dialogue:
Minority of pauses at TS (34% recall)
Most of TS is marked as a pause (57% precision)