The Development of Estonian Natural Language Dialogue Systems: A Case Study

Margus Treumuth
A dialogue system is a computational application by means of which a human interacts with a system using natural language.

- Natural and flexible access to digital information
- Decrease of routine work (online services, timetables, FAQ)
Goal of the project

• Build a dialogue system that operates in a constrained linguistic domain – theater information.
• The system should accept either typewritten or spoken language.
• The system should produce either typewritten or spoken output.
• The language used by the dialogue system is Estonian.
Project members and tasks

• Joint research project of the University of Tartu and the Tallinn University of Technology

• Dialogue management module – Tartu

• Speech recognition – Tallinn
  – Tanel Alumäe
  – Einar Meister

• Speech synthesis – Tallinn
  – Einar Meister
Dialogue system components

- Speech Recognition Module
- Dialogue Management Module
  - Morphological Analyzer and Generator
  - Query Generation and Date Recognition
  - Theater Information Database
- Speech Synthesis Module
Speech recognition

- segments the input stream into utterances
- produces a recognition hypothesis for each segment
- triggers barge-in, if it detects speech that continues for a configurable amount of time
- barge-in sends a signal to the speech synthesis module to stop any speech output
Dialogue manager - Tasks

• Initiates, maintains and records interaction
• Reaction to input (ask, inform, confirm, correct)
• Maintenance of dialogue history and context
• Interfaces to other system components
Dialogue manager - Components

- Morphological analysis
  - gives us base forms
- Date recognition
  - „next Sunday“ gives us 14.05.2006
- Play, Theater, City recognition
- Query Generation
  - fills the fields in a frame to send a query to the database
- Database
- Morphological synthesis
- Answer generation
Query generation

Spoken or typewritten input (in Estonian):
„I’d like to see Cats on next Sunday.“

<table>
<thead>
<tr>
<th>Field</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Date</td>
<td>?</td>
</tr>
<tr>
<td>Theater</td>
<td>?</td>
</tr>
<tr>
<td>Play</td>
<td>?</td>
</tr>
<tr>
<td>City</td>
<td>?</td>
</tr>
</tbody>
</table>
SELECT  t.theater_name,
       p.play_date,
       p.play_time,
       p.play_title
FROM   plays p, theaters t
WHERE  t.id = p.theater_id
       AND p.play_date >= CURDATE()
       AND p.play_date = STR_TO_DATE('14.05.2006', '%d.%m.%Y')
       AND p.play_title = 'Cats'
ORDER BY p.play_date
Speech synthesis

- two modalities
  - in text form
  - via speech output
- linguistic analysis of the input text
- orthographic text is converted into phonemic representation
- prosody model calculates the phoneme durations and the contour of fundamental frequency
- MBROLA model (Dutoit, et al., 1993)
The near future

• to produce better speech output, we should send some specific tags to the speech synthesis that include information about dialogue structure

• we should expand the knowledge base and probably switch to AIML (artificial intelligence markup language)