From Reaction to Prediction
Experiments with Computational Models of Turn-Taking

David Schlangen
Potsdam / Berlin Conversation Studies Lab
Department of Linguistics, University of Potsdam
dasl@ling.uni-potsdam.de

Background

Question: What distinguishes places in a conversation where a transition is possible from those where it isn't?
More specifically: What distinguishes them in terms of syntax and prosody?
Method: quantitative analysis; ML; how well can we predict turn-taking decisions?
Context: P/B-CSL devises and implements models of conversational competence.

Experiments

"is this x turn-final or not?"
- E1: utterances; human subj
- E2: utterances; ML
- E3: utterances, real data; ML
- E4: inter-pausal units; ML
- E5: all words; ML

Experiments

"is this x turn-final or not?"
- E1: utterances; human subj
- E2: utterances; ML
- E3: utterances, real data; ML
- E4: inter-pausal units; ML
- E5: all words; ML

Experiments

"is this x turn-final or not?"
- E1: utterances; human subj
data: created conversational situation where same (declarative) sentence would appear both turn-medial and turn-final. 8 sents; 3 spkrs. In German.
- E2: utterances; ML
- E3: utterances, real data; ML

Experiments

"is this x turn-final or not?"
- E1: utterances; human subj
data: 8 sentences, all recorded in turn-medial and turn-final position.
classifier: 24 university students
- E2: utterances; ML
- E3: utterances, real data; ML
Experiments

"is this x turn-final or not?"

- E1: utterances; human subj
  data: 8 sentences, all recorded in turn-medial and turn-final position.
  classifier: 24 university students
  question: will the speaker continue after this sentence or not? (continue y/n = wait/take)

- E2: utterances; ML
- E3: utterances, real data; ML

Results:

(Cutler & Pearson 1986) < us < (Barkhuysen et al. 2006)

- E2: utterances; ML
- E3: utterances, real data; ML

Experiments

"is this x turn-final or not?"

- E1: utterances; human subj
  data: 8 sentences, all recorded in turn-medial and turn-final position.
  classifier: 24 university students
  question: will the speaker continue after this sentence or not? (continue y/n = wait/take)

- E2: utterances; ML
- E3: utterances, real data; ML

- E4: inter-pausal units; ML
- E5: all words; ML

Experiments

"is this x turn-final or not?"

- E1: utterances; human subj
  data: same
  classifier: machine learners

- E2: utterances, real data; ML
- E4: inter-pausal units; ML
- Machine Classifiers

All ML experiments conducted with WEKA toolkit (Witten & Frank 2005)

Input represented by set of syntactic & prosodic features, computed for each word.

Evaluated using 10-fold cross-validation.

---

- Extracted Features

(cf. (Shriberg et al. 2000))

---

Experiments

"is this x turn-final or not?"

- E2: utterances; ML
  - data: pairs of recordings of sentence uttered in turn-medial and turn-final position, 8 sents.
  - classifier: machine learners

- E3: utterances, real data; ML
- E4: inter-pausal units; ML

---

Experiments

"is this x turn-final or not?"

- E2: utterances; ML
  - results:

<table>
<thead>
<tr>
<th></th>
<th>Crv</th>
<th>EI</th>
<th>Fse</th>
<th>CS</th>
<th>IC</th>
<th>Ch</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>MIP</td>
<td>68.9</td>
<td>19.3</td>
<td>5.6</td>
<td>0.64</td>
<td>0.02</td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIP</td>
<td>62.2</td>
<td>54.8</td>
<td>0.09</td>
<td>0.59</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MIP</td>
<td>60.4</td>
<td>92.2</td>
<td>0.03</td>
<td>0.50</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

- E3: utterances, real data; ML
- E4: inter-pausal units; ML

---

Experiments

"is this x turn-final or not?"

- E1: utterances; human subj
- E2: utterances; ML
- E3: utterances, real data; ML
- E4: inter-pausal units; ML
- E5: all words; ML

---

Experiments

"is this x turn-final or not?"

- E1: utterances; human subj
- E2: utterances; ML
- E3: utterances, real data; ML
- E4: inter-pausal units; ML
- E5: all words; ML
Experiments

"is this x turn-final or not?"

- E3: utterances, real data; ML
  - data: 100 dialogues from switchboard corpus;
    utterances (slash units)
  - classifier: machine learners

- E4: inter-pausal units; ML
- E5: all words; ML

Experiments

"is this x turn-final or not?"

- E3: utterances, real data; ML
- E4: inter-pausal units; ML
- E5: all words; ML

Experiments

"is this x turn-final or not?"

- E1: utterances; human subj
- E2: utterances; ML
- E3: utterances, real data; ML
- E4: inter-pausal units; ML
- E5: all words; ML

Experiments

"is this x turn-final or not?"

- E4: inter-pausal units; ML
- E5: all words; ML

Experiments

"is this x turn-final or not?"

- E3: utterances, real data; ML
- E4: inter-pausal units; ML
- E5: all words; ML

Data: swbd; units delineated by pauses
(varying the minimal length)

Task: take v wait (≈ turn-end sil. v hesitation)
Classifier: machine learners
(cf. e.g. (Ferrer et al. 2002, 2003))
Experiments

"is this x turn-final or not?"

- E3: utterances, real data; ML
- E4: inter-pausal units; ML
- E5: all words; ML

<table>
<thead>
<tr>
<th></th>
<th>ChF</th>
<th>PhF</th>
<th>C%</th>
<th>E%</th>
<th>Cx</th>
<th>E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mo</td>
<td>92</td>
<td>9.6</td>
<td>4.3</td>
<td>6.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fs</td>
<td></td>
<td>6.3</td>
<td>6.3</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Experiments

"is this x turn-final or not?"

- E3: utterances, real data; ML
- E4: inter-pausal units; ML
- E5: all words; ML
- E4: inter-pausal units; ML

data: all words in switchboard task: is (the utterance up to) this word turn-final or not?

Experiments

"is this x turn-final or not?"

- E3: utterances, real data; ML
- E4: inter-pausal units; ML
- E5: all words; ML
- E5: all words; ML

Results:

- Perf. worse; same int/f0 pattern
- Hard task; main predictive power: syntax

Summary / Conclusions

"is this x turn-final or not?"

- E1: utterances; human subj
  - Better than chance; better at wait
- E2: utterances; ML
  - Better at wait; int/f0 same for take; f0 for wait
- E3: utterances, real data; ML
  - Perf. worse; same int/f0 pattern
- E4: inter-pausal units; ML
  - Prosodic info improves over baseline
- E5: all words; ML
  - Hard task; main predictive power: syntax
Conclusions

*Question was: what distinguishes.*
*only some way towards answering this, more: are they distinguished?*
*E5 task is of course a poor approximation of real task of predicting TRPs: this is not done on a word-by-word basis.*

Future Work

*incremental parsing for prediction; realisation of filter model; prosody has veto*
*move further backwards: predict end in x ms*
*fixed window rather than word; using ASR results, etc.*

Thank you for your attention!

Acknowledgements:
thanks to the Potsdam Dialogue Group for discussion!

Diagrams produced with Zeitwort:
http://www.sourceforge.net/projects/zeitwort

--- this is the Berlin TV tower, in case you were wondering. Shown for no particular reason.

Corpus for E1 & E2

Instructions to speakers: “There are two sets of cards, one with situations described in sentences, the other with pictures of these situations. Your task is to read out the descriptions so that B [a confederate] can identify the cards.”

Instruction to B: give backchannels, some task related chat.