Does Information Theory constrain the usage of fragments? – An experimental study
Robin Lemke  Lisa Schäfer  Ingo Reich

Background: Fragments and Information Theory

Fragment (following Morgan, 1973)

Antecedentless overtly nonsentential utterance which expresses propositional content and has illocutionary force

(1) a. [Passenger to taxi driver:] “To the university, please.”
   b. [Ticket inspector to passenger:] “Your ticket, please.”

When are fragments licensed?

- Delete up to recoverability (Barton, 1998)
- When a salient antecedent is available (Stainton, 2006)
- When there is a high probability for the hearer to correctly interpret it (Bergen and Goodman, 2014)

Uniform Information Density (Levy and Jaeger, 2007))

- Information (or Surprisal, Hale 2001) of a word can be measured as $S = -\log p(\text{word}|\text{context})$
- Speakers intend to communicate at a uniform rate without exceeding the capacity of the communicative channel

Estimating event probabilities from a script corpus

DeScript corpus (Wanzare et al., 2016)

- Crowd-sourced corpus of script knowledge
- 100 individual descriptions for 40 scenarios
- Description in list form, many elliptical utterances

Method

- Parse corpus with Stanford parser (Klein and Manning, 2003)
- Extract events (verb + noun, Manshadi et al., (2008))
- Manually annotate ellipses and disambiguate pronouns
- Manually unify synonym events
- Calculate bigram language models (SRILM, Stolcke, (2002))

Sample item (original in German)

(2) Today Annika and Jenny want to cook a large serving of pasta. Annika put a pot with water on the stove. Then she turned the stove on. After a few minutes, the water started to boil. Now Annika says to Jenny:
   a. Pour the pasta into the water, please! Predictable
   b. Set the table, please! Unpredictable
   c. The pasta, please! Predictable
   d. The table, please! Unpredictable

Discussion and next steps

- As expected, predictable fragments are significantly better than unpredictable ones, but even those are worse than full sentences → Politeness constraint?
- The more script knowledge subjects have, the more they differentiate between predictable and unpredictable fragments
- Production study: Gather naturalistic data on what people would say in described situations
- Rating study with production data: Are specifically the informative, entropy-reducing, parts of the utterance maintained?

Selected References