The Effect of Correcting Grammatical Errors on Parse Probabilities

Joachim Wagner and Jennifer Foster

National Centre for Language Technology
School of Computing
Dublin City University
Ireland

IWPT 2009
Paris
8th October 2009
Parsing and Grammaticality Judgements

Probabilistic treebank-derived grammars
- Overgenerate / parse almost any input
- Architecture for precision grammars not applicable

Can such grammars be used to judge a sentence’s grammaticality?
Parse Probability

Can this number tell us anything about grammaticality?

Other factors which affect parse probability

1. Sentence length
2. Lexical choice
Experiment

Do grammatical errors affect parse probability in a negative way?

Procedure

1. Three parallel error corpora
2. Parse with a state-of-the-art head-lexicalised generative probabilistic parser
3. Examine the difference between the parse probabilities of the most likely trees for each grammatical/ungrammatical pair
Experiment

Three parallel error corpora

1. **Foster**: 1,132 pairs containing native and non-native English
2. **Gonzaga**: 500 transcribed spoken utterance pairs produced by learners in a classroom setting
3. **BNC**: common grammatical errors automatically introduced into well-formed sentences

Benefits

- Parallelism: control for length and lexical choice
- Artificial errors: more data & effect of particular error types can be examined
Foster Results

(native and non-native English, authentic errors)
Gonzaga Results
(non-native learner English, spoken, authentic errors)
BNC Results

(native English, artificial errors)
Results by Error Type

- Real-word spelling errors more likely to lower probability than agreement and verb form errors
  Anyway, the/they left us alone. -76/-65

- Agreement errors involving an article most likely to have negative effect
  Yeah that’s an ideas/idea -66/-52

- Missing word errors often increase the probability
  Doreen ε/sounded incredulous. -64/-71
  except for missing function words
  I ε/’d be only too happy to help. -68/-65
Summary

Observation

- Grammatical errors do tend to have a negative effect on parse probability
- Supports evidence from psycholinguistic research that grammaticality and frequency are linked. (Crocker and Keller 2006)
- Error detection systems: adding it as a feature does not appear to help a lot (Sun et al. 2007)
  - Controlling for the other factors is vital - this is ongoing work!
Thanks to ...

the reviewers

ICHEC
Irish Centre for High-End Computing

IRCS
Irish Research Council for Science, Engineering & Technology
Investing in People and Ideas

Josef van Genabith

BRITISH NATIONAL CORPUS

You!