Evaluating and Extending the Coverage of HPSG Grammars: A Case Study for German

Jeremy Nicholson, Valia Kordoni, Yi Zhang, Timothy Baldwin, Rebecca Dridan

Department of Computational Linguistics
Saarland University
& DFKI GmbH

Department of Computer Science and Software Engineering
& NICTA Victoria Research Labs
University of Melbourne

28 May 2008

Deep Lexical Grammars

- Deep grammars provide a full analysis, more semantic information than shallower tools
- Tendency to emphasise precision over recall can cause poor coverage
- HPSG grammars, parsing tools from the DELPH-IN initiative
- Our aim: take a "snapshot" of the grammar, examine potential for expansion

Analysis of a "Broad–Coverage" Grammar

- "Beauty and the Beast" (2004):
  - Use the ERG to parse 20K sentences from the BNC
  - Analyse sources of parse failures
- "Evaluating and Extending" (Today):
  - Use GG to parse 612K sentences from Frankfurter Rundschau
  - Evaluate errors over 1K sentences
  - Use lexical type prediction to increase coverage

Corpus Analysis of the Grammar

- Ran a large grammar out–of–the–box on a very different corpus
- Lexical span: ERG - 32%; GG - 28%
- Sentences with correct reading attested: ERG - 83%; GG - 85%

	No span	Span, no parse	≥ 1 parse
ERG	68%	14%	18%
GG	72%	16%	12%

Lexical Gaps for GG

Lexical gaps:

Error Type	Proportion
lexical entries	33%
proper nouns	22%
noun compounds	30%
tokenisation	12%
garbage strings	2%

Lexical Gaps for GG

Lexical entry:
Aufgrund des ruhigeren Geschäftsverlaufs rechnet Maier für 1992 mit einem "leicht rückläufigen" Ergebnis.

Noun compound:
Das Türelement läßt sich hinter die Verkleidung schieben und wird damit unsichtbar.

Sophisticated tokenisation could account for proper nouns, noun compounds, tokenisation errors.

Parsing Errors for GG

Parsing Errors for GG:

Error type	Proportion
constructional gap	39%
lexical item gap	47%
multi–word expression	7%
spelling	4%
fragment	3%

Parsing Errors for GG

Constructional gap:
BREMEN, 4. Februar.

Lexical item gap:
Beginn ist um 19 Uhr in der Stadthalle.

Multi–word expression:
Der Opfer dieser Verbrechen der Nationalsozialisten gedachte die Stadt Bad Homburg gestern abend.

Similar distribution observed in Beauty and the Beast.

Lexical Acquisition

- Baldwin (2005): use a range of morphological, syntactic, semantic features for predicting lexical type class of unknown token/type
- e.g. *Katze* in *Die Katze ist schwarz.* is one of `count-noun-le`, `mass-noun-le`, `count-noun-mass-unit-le`, `deverbal-noun-le`...

Lexical Acquisition

- Feature set from Zhang and Kordoni (2006): prefixes/suffixes, 2 tokens of context, 2 types of context
- Token–wise prediction on the GG treebank (MaxEnt, cross–validation)
- Limit evaluation to "unknown words" (type–wise)
- Accuracy approaches 60%

Lexicon Extension

- Using the MaxEnt model from the treebank, predict lexical types for unknown tokens within Frankfurter Rundschau
- Intrinsic evaluation not possible
- Thresholding MaxEnt at 10% likelihood, add 1130 lexemes to the lexicon
- Further 9% coverage, 83% of these had at least one parse

Summary

- Change of 12% of parsed sentences at 85% precision to about 20% at 84% precision
- This means getting more "easy" sentences
- Scope for improving the grammar, parsing strategy (shallow methods to improve deep parsing)