Precision Grammar Implementation for Linguistic Hypothesis Testing

Emily M. Bender
University of Washington
ebender@u.washington.edu

Dan Flickinger
CSLI Stanford
danf@csli.stanford.edu

Stephan Oepen
Universitetet i Oslo & CSLI Stanford
oe@csli.stanford.edu

http://lingo.stanford.edu/courses/07/lsa/
Every time I fire a linguist, system performance goes up.

[Fred Jelinek, 1980s]
Back Then: The Rationalist vs. Empiricist Stand-Off

Every time I fire a linguist, system performance goes up.

[Fred Jelinek, 1980s]

Competition of Paradigms

- Rationalist: formally encode linguistic and extra-linguistic knowledge;
- empiricist: statistical models approximate human language competence;
- Jelinek eventually turned off the lights — LFG & HPSG groups stable;
  → keep focus: ‘deep’ linguistic approaches required for long-term success.
Competing Approaches (1 of 2)

Can you send me copies of all checks in December?

Statistical Part-of-Speech Tagging (96.7% Accuracy)

<table>
<thead>
<tr>
<th></th>
<th>1.0</th>
<th>1.0</th>
<th>0.98</th>
<th>1.0</th>
<th>1.0</th>
<th>1.0</th>
<th>1.0</th>
<th>1.0</th>
<th>.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MD</td>
<td>PRP</td>
<td>VB</td>
<td>PRP</td>
<td>NNS</td>
<td>IN</td>
<td>DT</td>
<td>NNS</td>
<td>IN</td>
</tr>
</tbody>
</table>
| Can you send me copies of all checks in December ?

Text Classification (~85% Accuracy)

CheckCopyRequest 0.6934, CheckBookRequest 0.0247, StatementCopyRequest 0.0066, ...
Competing Approaches (2 of 2)

\[
\langle h_1, \\
\{ h_1: \text{int}_m(h_2), \ h_3: \text{can}_v\text{modal}(e_4, h_5), \ h_7: \text{send}_v(e_8, x_9, x_{10}, x_{11}), \\
\ h_{12}: \text{pronoun}_q(x_9, h_{13}, h_{14}), \ h_{15}: \text{pron}(x_9 \{ 2nd \}), \\
\ h_{16}: \text{pronoun}_q(x_{10}, h_{17}, h_{18}), \ h_{19}: \text{pron}(x_{10} \{ 1sg \}), \\
\ h_{20}: \text{bare}_q(x_{11}, h_{21}, h_{22}), \ h_{23}: \text{copy}_n\text{of}(x_{11} \{ \text{pl} \}, x_{24}), \\
\ h_{25}: \text{all}_q(x_{24}, h_{26}, h_{27}), \ h_{28}: \text{check}_n(x_{24} \{ \text{pl} \}), \\
\ h_{28}: \text{temp}_\text{loc}(\_, x_{24}, x_{29}), \ h_{30}: \text{proper}_q(x_{29}, h_{31}, h_{32}), \ h_{33}: \text{mofy}(x_{29}, "DEC") \}, \\
\{ h_2 = q h_3, \ h_5 = q h_7, \ h_{13} = q h_{15}, \ h_{17} = q h_{19}, \ h_{21} = q h_{23}, \ h_{26} = q h_{28}, \ h_{31} = q h_{33} \} \rangle
\]

(Truth-Conditional or) Logical-Form Semantics

+ high-level abstraction; grounded in entities and relations \(\rightarrow\) inference;
– very difficult to construct (correctly, with broad-coverage) and process.
Ambiguity Resolution Remains a (Major) Challenge

The Problem

- With broad-coverage grammars, even moderately complex sentences typically have multiple analyses (tens or hundreds, rarely thousands);
- unlike in grammar writing, exhaustive parsing is useless for applications;
- identifying the ‘right’ (i.e. intended) analysis is a very hard problem (AI);
- inclusion of (non-grammatical) sortal constraints is generally undesirable.

Candidate Approaches

- Heuristic scoring rules applied to (classes of) lexical items and rules;
- ‘optimality’ projection: accumulate quality marks and rank globally;
- now dominant: probabilistic models for on- or off-line parse selection.
LinGO Redwoods

— A Rich and Dynamic Treebank for HPSG —

Stephan Oepen, Daniel P. Flickinger,
Kristina Toutanova, Christopher D. Manning

Center for the Study of Language and Information
Stanford University

oe@csli.stanford.edu
LinGO Redwoods: a Rich and Dynamic Treebank

- Tie treebank development to existing broad-coverage grammar;
- hand-select (or reject) intended analyses from parsed corpus;
- [Carter, 1997]: annotation by basic discriminating properties;
- record annotator decisions (and entailment) as first-class data;
- provide toolkits for dynamic mappings into various formats;
- semi-automatically update treebank as the grammar evolves;
- integrate treebank maintenance with grammar regression testing.
Annotation: Basic Discriminating Properties

- Extract minimal set of *basic discriminants* from set of HPSG analyses;
- typically easy to judge, need little expert knowledge about grammar;
- allow quick navigation through parse forest and incremental reduction;
- *constituents* use of particular construction over substring of input;
- *lexical items* use of particular lexical entry for input token;
- *labeling* assignment of particular abbreviatory label to a constituent;
- *semantics* appearance of particular key relation on constituent;
- Stanford undergraduate annotates some 2000 sentences per week.

- Regularly propagate discriminants into new version of parsed corpus;
Redwoods Representations: Native Encoding

```
yesno
    | hcomp
  |   
sailr you
|   |    
do1_pos you
|   |    
do

hcomp

hcomp

hcomp

hcomp

hcomp

hcomp

bse_verb

bse_verb

bse_verb

verb

toc_prop

hadj_uns

to

meet_v1

meet

meet

on_day

proper_np

on

noptcomp

sing_noun

tuesday1

Tuesday
```
Derived Encodings: Labeled Phrase Structure Trees

- reconstruct full HPSG analysis from derivation tree;

- match underspecified feature structure ‘templates’ against each node;

- optionally, collapse or suppress nodes.

\[
\text{SYNSEM.LOCAL.CAT} = \begin{cases} \text{label} \\ \text{HEAD \ verbal} \\ \text{VAL} \ \text{SUBJ} \ \langle \rangle \\ \text{COMPS} *\text{olist}* \end{cases} \equiv \text{‘S’}
\]
Probabilistic Context-Free Grammars (PCFGs)

<table>
<thead>
<tr>
<th>probability</th>
<th>context-free rule (i.e. local tree of depth one)</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/3 = 0.66</td>
<td>HEAD-SPECIFIER → HEAD-COMPLEMENT-0 HEAD-COMPLEMENT-0</td>
</tr>
<tr>
<td>1/3 = 0.33</td>
<td>HEAD-SPECIFIER → HEAD-SPECIFIER HEAD-COMPLEMENT-1</td>
</tr>
<tr>
<td>1/1 = 1.00</td>
<td>HEAD-COMPLEMENT-1 → VERB-WORD-TRANSITIVE HEAD-SPECIFIER</td>
</tr>
<tr>
<td>2/4 = 0.50</td>
<td>HEAD-COMPLEMENT-0 → NOUN-WORD</td>
</tr>
<tr>
<td>2/4 = 0.50</td>
<td>HEAD-COMPLEMENT-0 → DET-WORD</td>
</tr>
</tbody>
</table>
Parse Selection: The Maximum Entropy School

Conditional Parse Selection

• Local independence assumption is not true for unification grammars;
• PCFG unable to ‘learn’ from negative data, e.g. dis-preferred parses;
→ conditional model: given some context, sample properties of events.

Conditional Parse Selection

Given a sentence $s$ and a set of trees $\{t_1 \ldots t_n\}$ assigned to $s$ by some grammar, find the tree $t_i$ that maximizes $p(t_i|s)$. Assuming a set of features $\{f_1 \ldots f_m\}$ with corresponding weights $\{\lambda_1 \ldots \lambda_m\}$, the conditional probability for tree $t_i$ is given by:

$$p(t_i|s) = \frac{\exp \sum_j \lambda_j f_j(t_i)}{\sum_{k=1\ldots n} \exp \sum_j \lambda_j f_j(t_k)}$$

(1)
Redwoods Applications: Parse Disambiguation

- Manning & Toutanova (Stanford): generative and conditional models;
- Baldridge & Osborne (Edinburgh): active learning and co-training;
- Fujita, Bond, et al. (NTT): semantics and ontologies in parse selection;
- feature selection: phrase structure, morpho-syntax, dependencies;
- ten-fold cross validation: score against annotated gold standard;
- preliminary results: 80% exact match parse selection accuracy;
- on-line use in parser: n-best beam search guided by MaxEnt scores;
- alternatively, full parse forest (polynomial) plus selective unpacking.
Conclusions — Background Material

- ‘Deep’ grammar-based processing requires adequate stochastic models;
- basic research needed on acquisition and application of stochastic models;
- no existing treebank resources with suitable granularity and flexibility;
- LinGO Redwoods treebank based on existing open-source technology;
- tied to broad-coverage HPSG grammar: advantages and disadvantages;
- rich in available information, dynamic in data extraction and evolution.

Grammar and Treebank available from: http://redwoods.stanford.edu/
Based on Research and Contributions of

Tim Baldwin, John Beavers, Ezra Callahan
Emily M. Bender, Kathryn Campbell-Kibler,
John Carroll, Ann Copestake,
Rob Malouf, Ivan A. Sag,
Stuart Shieber, Tom Wasow,
and others.