The Matrix: Future Directions
Wrap up

Ling 567
March 9, 2020
Overview

• Wrap up/reflections

• Matrix: Future directions

• AGGREGATION
Goals: Of Grammar Engineering

• Build useful, usable resources
• Test linguistic hypotheses
• Represent grammaticality/minimize ambiguity
• Build modular systems: maintenance, reuse
Goals: Of this course

• Mastery of tfs formalism

• Hands-on experience with grammar engineering

• A different perspective on natural language syntax

• Practice building (and debugging!) extensible system

• Contribute to on-going research in multilingual grammar engineering
Big take-aways this term for me

• Automatically inferred choices files are still too noisy

• eng2yaq presents a case where the VPM framework possibly breaks down

• append-lists for long distance dependencies don’t get along (as implemented) with coordination

• … lots of specific bugs noted!
Reflections

• Where have the analyses provided by the Matrix (or suggested by the labs) seemed like a good fit?

• Where have they been awkward?

• What have you learned in this class about syntax?

• ... about knowledge engineering for NLP?

• ... about computational linguistics in general?

• ... about linguistics in general?

• What did working with a test corpus show you about the process of scaling to real-world text?
Feedback: Individual projects

• Was the workload reasonable?

• How would you have taken advantage of working with a partner?

• What was better about working alone?
More reflections

- Semantic representations are important
  - It’s easier to work on them if they serve as an interface to something

- Analyses of phenomena interact
  - The more streamlined/motivated the analysis of each phenomenon is, the smoother the interactions

- What interactions did you encounter?
More reflections: model and modeling domain

• From 566: Distinction between the model (HPSG grammar fragment) and the modeling domain (there: English).

• How did this play out in 567?
Future directions overview

- More libraries (and semantic harmonization)
- How this class might evolve
- AGGREGATION
More libraries

• In progress: Wh questions
• Next up?
  • Pronouns, demonstratives, COG-ST
  • Other non-verbal predicates
  • Other intersective modifiers
  • Numeral classifiers
  • More verb subcategorization
  • Information structure in wh questions
  • Information structure in free word order languages
Creating a library for the customization system

- Choose phenomenon
- Review typological literature on phenomenon
- Refine definition of phenomenon
- Conceptualize range of variation within phenomenon
- Review HPSG (& broader syntactic) literature on phenomenon
- Pin down target MRSs
- Develop HPSG analyses for each variant
- Implement analyses in tdl
- Develop questionnaire
- Extend python backend
- Run regression tests
- Test with pseudo-languages
- Test with illustrative languages
- Test with held-out languages
- Add tests to regression tests
- Add to MatrixDoc pages
How to evaluate a library

- Pseudo-languages
- Illustrative languages
- Held-out languages
- Test suites
- Choices files
- Error analysis
More libraries/reflection from current class

• What do you most wish was available in the customization system, based on what came up in your test suite?

• In your test corpus?
Evolution of 567

• New phenomena: Wh-questions, possessives, relative clauses, while-clauses ...?

• Ever bigger jump start --- reaching the limit on this one

• Time to “clean up” noisy inferred morphology early on

• How did these work out?:
  • Partnership with field linguists
  • Work with linguist-provided corpora

• Coverage-driven labs seem most satisfying (MT demo, corpus coverage). Is this true? Can the course be rebalanced to do more of this?
AGGREGATION Project:
Motivation & overview

• Precision grammars are potentially useful for endangered language documentation (Bender et al 2012)

• Field linguists produce extremely rich annotations in the form of interlinear glossed text

• The Grammar Matrix provides a mapping from grammar specifications to precision grammars

• Can we infer sufficiently accurate and complete grammar specifications from IGT?

- Interlinear glossed text (IGT) is an extremely rich data type

- IGT exists in plentiful quantities on the web, even for low resource languages

- Example from Chintang [ctn]:

  akka ita    khurehê
I bought a pair of shoe.

(IGT from Bickel et al 2012)
William Carlos Williams was right: "wheelbarrow" has a whopping 5 dependents!

so much depends upon a red wheelbarrow glazed with rain water beside the white chickens
Bender et al 2013: Inferring large-scale properties

Task 1: Major constituent word order

- Count word order patterns in projected trees
- Calculate ratios of OS:SO etc
- Plot points for each language in 3D space
- Compare to hypothesized canonical points for each word order
- V2 (and not free) if SVO, OVS >> SOV, OSV

Figure 2: Three axes of basic word order and the positions of canonical word orders.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Inferred WO</th>
<th>Baseline</th>
</tr>
</thead>
<tbody>
<tr>
<td>DEV1</td>
<td>0.900</td>
<td>0.200</td>
</tr>
<tr>
<td>DEV2</td>
<td>0.500</td>
<td>0.100</td>
</tr>
<tr>
<td>TEST</td>
<td>0.727</td>
<td>0.091</td>
</tr>
</tbody>
</table>

Table 2: Accuracy of word-order inference

@emilymbender

- General parameters like word order alone won’t lead to a usable grammar
- Also required: lexicon and morphotactics (and morphophonology…)
  - Create lexical rules for each morpheme, with associated form and morphosyntactic and morphosemantic features
  - Group morphemes into position classes
  - Determine ordering relations
  - Lexicon: part of speech, case frame, argument optionality…
Lepp et al 2019: Visualizing inferred morphotactics
Lepp et al 2019: Visualizing inferred morphotactics
End-to-end evaluation with Chintang [ctn] (Zamareva et al 2019)

<table>
<thead>
<tr>
<th>Choices file</th>
<th># verb entries</th>
<th># noun entries</th>
<th># verb affixes</th>
<th># noun affixes</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORACLE</td>
<td>899</td>
<td>4750</td>
<td>233</td>
<td>36</td>
</tr>
<tr>
<td>BASELINE</td>
<td>3005</td>
<td>1719</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>FF-AUTO-GRAM</td>
<td>739</td>
<td>1724</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>MOM-DEFAULT-NONE</td>
<td>1177</td>
<td>1719</td>
<td>262</td>
<td>0</td>
</tr>
<tr>
<td>INTEGRATED</td>
<td>911</td>
<td>1755</td>
<td>220</td>
<td>76</td>
</tr>
</tbody>
</table>

Table 3: Amount of lexical information in each choices file

<table>
<thead>
<tr>
<th>choices file</th>
<th>lexical coverage (%)</th>
<th>parsed (%)</th>
<th>correct (%)</th>
<th>readings</th>
</tr>
</thead>
<tbody>
<tr>
<td>ORACLE</td>
<td>116 (12.5)</td>
<td>20 (2.2)</td>
<td>10 (1.1)</td>
<td>1.35</td>
</tr>
<tr>
<td>BASELINE *</td>
<td>38 (0.4)</td>
<td>15 (1.6)</td>
<td>8 (0.9)</td>
<td>27.67</td>
</tr>
<tr>
<td>FF-AUTO-GRAM</td>
<td>18 (1.9)</td>
<td>4 (0.4)</td>
<td>2 (0.2)</td>
<td>5.00</td>
</tr>
<tr>
<td>MOM-DEFAULT-NONE</td>
<td>39 (4.2)</td>
<td>16 (1.7)</td>
<td>3 (0.3)</td>
<td>10.81</td>
</tr>
<tr>
<td>INTEGRATED</td>
<td>105 (11.3)</td>
<td>32 (3.4)</td>
<td>15 (1.6)</td>
<td>91.56</td>
</tr>
</tbody>
</table>

* We report slightly different results for lexical coverage and average readings for the baseline than Bender et al. (2014) because we removed determiners from the choices file.

Table 4: Results on 930 held-out sentences
Extending inference (Howell in progress)

- Previously available: major constituent word order, case systems, case frames for verbs, case values for nouns

- Adding: argument optionality, coordination, PNG on nouns and agreeing categories, tense/aspect/mood, sentential negation, auxiliaries, case-marking adpositions

- Initial system tested in Ling 567 as starting grammar specifications (noisy!)

- Testing on 15 languages: 10 dev, 5 held-out
  - Coverage, ambiguity, treebanked accuracy
Extending inference (Howell in progress)

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Development1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Abui [abz]</td>
<td>1569</td>
<td>54.75%</td>
<td>43.98%</td>
<td>8.28%*</td>
<td>7.01%*</td>
<td>2365</td>
</tr>
<tr>
<td>Chintang [ctn]</td>
<td>9785</td>
<td>22.24%</td>
<td>12.38%</td>
<td>2.15%*</td>
<td>1.02%*</td>
<td>5648</td>
</tr>
<tr>
<td>Nuuchahnuht [nuk]</td>
<td>641</td>
<td>6.08%</td>
<td>3.43%</td>
<td>1.09%</td>
<td>1.09%</td>
<td>18</td>
</tr>
<tr>
<td>Matsigenka [mcb]</td>
<td>349</td>
<td>20.59%</td>
<td>8.82%</td>
<td>0.57%</td>
<td>0.57%</td>
<td>6827</td>
</tr>
<tr>
<td>Wambaya [wmb]</td>
<td>818</td>
<td>13.81%</td>
<td>3.91%</td>
<td>2.32%</td>
<td>0.00%</td>
<td>6</td>
</tr>
<tr>
<td><strong>Development2</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Haiki [yaq]</td>
<td>2235</td>
<td>21.39%</td>
<td>12.39%</td>
<td></td>
<td></td>
<td>93</td>
</tr>
<tr>
<td>Lezgi [lez]</td>
<td>1197</td>
<td>12.87%</td>
<td>10.03%</td>
<td></td>
<td></td>
<td>19121</td>
</tr>
<tr>
<td>Meithei [mni]</td>
<td>1717</td>
<td>8.56%</td>
<td>7.92%</td>
<td></td>
<td></td>
<td>7139</td>
</tr>
<tr>
<td>Tsowa-Tush [bbl]</td>
<td>1611</td>
<td>30.11%</td>
<td>24.77%</td>
<td></td>
<td></td>
<td>2470</td>
</tr>
</tbody>
</table>

Table 1: Coverage and Ambiguity for Development Languages. Results are averages across 10 splits. * indicates results for only a single split.
External resources: WALS (Dryer & Haspelmath 2013)

• To what extent do the features in WALS map to Grammar Matrix grammar specifications? (Almeida et al 2019)
  • Some are mappable, but often without the degree of specificity required

• Where they do map, what is the best way to leverage them in inference of grammar specifications? (Zhang et al 2019)
  • Currently testing: use WALS info, when available, to constrain what inference looks for

• Could we do better with Autotyp? (Bickel & Nichols 2002)
Overview

• Wrap up/ reflections

• Matrix: Future directions

• Next time: MMT extravaganza and course evals