

Knowledge Engineering for NLP

March 5, 2007

Wrap-Up, Future Directions

Overview

- Transfer debugging
- Wrap-up/reflections
- Matrix: Future directions

Wrap-up/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?

Future Directions Overview

- More modules (and semantic harmonization)
- How this class will evolve
- Auto-generated transfer rules
- Lexical acquisition
- Ontological annotation
- Naive-user extensions to MT demo

More modules

- Lexical rules through customization interface
- Person, number, gender and verb-argument agreement
- Case
- Tense and aspect
- Demonstratives
- Argument optionality
- Extensions to negation and question modules

Evolution of 567

- New phenomena: copular sentences, wh-questions, tense and aspect
- Bigger jump start
- Partnership with field linguists
- Work with small corpora

Autogenerated transfer rules

- Identify “grammaticalized” differences in MRSs
- “Publish” choices along those dimensions for each grammar
- Create a library of transfer rules from property to property:
 - Pro-drop to pronouns (and vice versa)
 - Mismatches in demonstrative distinctions
 - Can \leftrightarrow the possibility exists

Lexical acquisition

- How can we import lexical entries from other linguistic sources (e.g., FIELD lexicons, ODIN)
- How big do the grammars have to be before we can embark on (semi-)automated lexical acquisition?
- How can we most effectively leverage human effort?
- How do we know when we're missing the appropriate type?

Ontological annotation

- Annotate grammars with links to GOLD
 - Locate which constraints contribute to which phenomena
 - Index analyses for discovery
- Annotations in Matrix core
- Annotations in customization script-generated tdl
- Support for user annotation

Naive User Extensions to MT Demo

- Add grammars just from configuration script (okay, still need a linguist here)
- Add vocabulary
- Structural disambiguation \Rightarrow treebank

And the distant future...

- Multiword expressions
- CALL for endangered languages
- ...

Overview

- Transfer debugging
- Wrap-up/reflections
- Matrix: Future directions