

Grammar Engineering

April 11, 2005

Test suites

Overview

- General announcements
- Evaluation in computational linguistics
- Uses of test suites
- Evaluating precision grammars
- Levels of adequacy
- Test suite software ([incr tsdb()] demo)
- More matrix tour

Ask more questions!!

- About: lab instructions, phenomena in your language, the Matrix, the LKB, HPSG
- Who's using the FAQ?
- Who spent > 10 minutes figuring out an LKB error?

Test suite cx by Wednesday?

- Case and agreement
- Argument optionality
- Modification
- Polar questions
- Imperatives
- Modals (*can*)
- Sentential negation
- Coordination: of NPs, VPs, Ss, other constituents

Also by Wednesday

- Vocabulary
- Case: overt case or not, ergative or accusative, cases on core arguments, where case is marked morphologically, any funnyness (case changes with definiteness, etc)
- Agreement: person, number, gender, definiteness; subject-verb, object-verb, determiner-noun, adjective-noun
- If your language has neither case nor agreement, talk to me about something else to code

... So: testsuites feasible or no?

Evaluation and computational linguistics

- Why is evaluation so prominent in computational linguistics?
- Why is it not so prominent in other subfields of linguistics?
- What about CS?

Uses of testsuites

- How far do I have left to go?
 - Internal metric
 - Objective comparison of different systems
 - How do you evaluate precision grammars?
- Where have I been?
 - Regression testing
 - Documentation

Kinds of test suites

- Hand constructed
 - Controlled vocabulary
 - Positive and negative examples
 - Controlled ambiguity
- Corpus based
 - More open vocabulary
 - Greater ambiguity
 - Haphazard ungrammatical examples
 - Application-focused
- Which kind for which use?

Evaluating precision grammars

- Coverage against a corpus
 - Which corpus?
 - Challenges of lexical acquisition
- Coverage of phenomena
 - How does one choose phenomena?
 - What did TSNLP do?
- Comparison across languages

Levels of adequacy (1/2)

- grammaticality
- “right” structure
- “right” dependencies
- exact match semantics

Levels of adequacy (2/2)

- Only legitimate parses (how do you tell?)
- Some set of parses including the preferred one
- Preferred parse only/first
- Preferred parse within first N

Test suite software

- What does the LKB batch parse utility do?
- What else would you like it to do?
- → [incr tsdb()] demo

Overview

- General announcements
- Evaluation in computational linguistics
- Uses of test suites
- Evaluating precision grammars
- Levels of adequacy
- Test suite software ([incr tsdb()] demo)
- More matrix tour