Grammar Engineering: TDL

05.2012

1. Download *grammar02* from `http://www.coli.uni-saarland.de/~yzhang/ge/grammar/grammar02.tgz`. Load the grammar and notice how GLB types are automatically generated.

2. Download *grammar03* from `http://http://www.coli.uni-saarland.de/~yzhang/ge/grammar/grammar03.tgz`. This is a very small grammar that demonstrates how a simple-minded context-free phrase structure grammar is implemented in LKB. Look through the files and understand how it works. Try batch-parse on test sentences in *"items"* and observe the results. Document your discoveries in the written report.

3. Extending the lexicon
   The grammar now has only 6 lexical entries: 2 determiners, 2 noun entries and 2 verb entries. You need to add lexical entries for the following words:

   - *that*: determiner
   - *cat*: countable noun, in both singular and pural forms (Careful! Avoid naming clash with existing type *cat* which is abbreviation for *category*).
   - *chase*: transitive verb, in base form and 3rd-person-singular form

   Reparse the test items and document the change of grammar coverage.

4. Extending the grammar rules
   You should have noticed that the grammar comes with only two rules that simulate the following CFG grammar:

   ```
   S -> NP VP
   NP -> DET N
   ```

- Rewrite the existing two rules using abbreviatory list notation, so that they are more readable

- Add a rule for verb phrases so that transitive verbs can be covered by the grammar. You are free to introduce extra features or make changes to types or lexicon if necessary. Document your analysis and the change of grammar coverage.

5. Adding agreements
   The grammar comes with overgeneration due to the lack of person-number agreements. Introduce an AGR attribute to the type *cat* and modify the rules to garantee determiner-noun and subject-verb agreements. Test your grammar coverage and accuracy and document your analyses.

6. Using morphological rules
   The lexicon of the grammar comes in a bad shape, in that different forms of the same word are associated with different lexical entries. Add necessary morphological rules to the file *"inflr.tdl"* and make changes to the lexicon, so that:

   - There is only one entry for each word in the lexicon
   - The value of the AGR feature is properly set by the morphological rules

   Document your analyses and the grammar performance change.