

HPSG-MWE Lab session for PARSEME Training School, Prague
Thursday, 22-01-2015, 9:00
Dan Flickinger

Getting started

You will need a web browser and a terminal window, both of which can be launched from the menu bar at the bottom of your screen. We'll start by taking a tour of an online treebank of about 1.5 million words of text parsed using the English Resource Grammar (ERG), to make some observations about MWEs.

MWEs in parsed corpus:

<http://moin.delph-in.net/ErgTop> : Linguistic Type Database

1. Inspect instances of these lexical types in a treebanked corpus:

=> Scroll down and click on each of the lexical types below to see lexical entries of these types, and to view treebanked examples of each. You may click on "More" for either the Lexical Examples or the Corpus Examples to see additional data.

Verb-particles ("We looked up the answer.")

v_p-np_le
v_p_le
v_p-pp_le

Determinerless PPs ("The book is on top of the bookcase.")

n_-_c-brno_le
n_-_c-brj*_le

=> Do any of these entries or analyses seem wrongly defined or annotated?

2. Inspect instances of MWE-specific syntactic rules in this treebank.

=> At the top of the page, click on "Rules" to get a list of syntactic rules in the treebank. Scroll down and click on the rule name below.

Adj-Noun+ed ("We have a long-eared rabbit at home.")

j_n-ed_c

MWEs in the ERG lexicon

3. Search in the ERG's lexicon file for word-with-spaces MWEs

=> Either open the following file in a text editor such as Emacs, or search it using a utility like 'grep' in the terminal window:

Lexicon file: `~parseme3/parseme/erg/lexicon.tdl`

=> Use this pattern, which only appears in words with multiple tokens in ORTH:
", (a double quote followed by comma)

=> How many of these entries are in the file?

4. Parsing individual sentences using online ERG demo

<http://lingo.stanford.edu/erg> : Demo

=> Parse the sentence:

We pointed out gently to them that he needs a ride.

=> What is the multi-word in this sentence?

=> Click the checkbox by "mrs" in the "output" options, and uncheck "eds", then click "Analyze" to reparse the sentence and produce MRS output.

5. Parsing to inspect and find MWEs

Setup:

```
PATH=/afs/ms/u/p/parseme3/parseme:$PATH
```

```
cd /afs/ms/u/p/parseme3/parseme
```

The parser we're using is ACE - see <http://moin.delph-in.net/AceTop>

Start ACE parser for interactive input

```
ace -g erg/erg.dat -lTf
```

=> Parse the sentence:

We pointed out gently that he needs a ride.

- right-click on V node just above "pointed" to see lex-entry feature structure
- type q to close that window
- right-click in the white space of the parse tree to get menu:
choose "Indexed MRS" to see semantics

6. Run ACE parser over a file (one sentence per line) and store just MRS

```
ace -g erg/erg.dat -lTf < corpus.txt > ~/out.txt
```

=> Count the number of occurrences of `_look_v_up_rel` (a semantic predicate)

```
ace -g erg/erg.dat -lf < corpus.txt > ~/out.txt, and store both deriv and MRS
```

=> Count the number of occurrences of `look_up_v1` (a lexical entry)