CG3 course
Constraint Grammar exercises

For the CG exercise section, we will use the new IDE for CG3, an integrated grammar development environment that you can install on your own computer, for both Windows, Mac and Linux, and which will allow you to write, run and debug grammars on either raw text or morphologically annotated cohort format. You can download the IDE, together with the newest version of the CG3 rule compiler, at:

http://goo.gl/mUOzWE

If for some reason this doesn't work for you, you can also avoid installation altogether by using the online VISL's CG lab interface, at http://beta.visl.sdu.dk/cglab.htm , linked from the general CG page at http://beta.visl.sdu.dk/constraint_grammar.html . In any case, CG lab is the old course page and has links to all course materials. Also, you can get morphological analyses for raw text for many languages from the CG lab interface, to copy into the IDE's input window.

Suggested reading: For doing the actual exercises, you don't have to be a Constraint Grammar expert, The introductury module should have provided you with some basics about the CG formalism as such, and the Vislcg/CG-3 grammar formalism in particular. Hopefully you'll also have a rough idea by now, from rule examples or Corpus examples, about the categories used in the VISL parsers. If you need help while doing the exercises, you can always check the details in the following references:

Vislcg formalism: http://beta.visl.sdu.dk/cg3_howto.html
Grammatical categories: http://beta.visl.sdu.dk/en/info/

For the exercises, we suggest that you copy the example grammar file (http://beta.visl.sdu.dk/cg_template_en for English, or for other languages http://beta.visl.sdu.dk/cg_template), and use it as a point of departure to add your own rules. Copy the file into the grammar window of the IDE or the CG lab, and change it there.

1. Preparation: Inspection

For a start, run the example grammar *as-is,* with a simple example sentence (a) or (if you feel confident enough) paragraph (b), and try to understand which rule does what. You can either do this in the IDE or command line on your own machine, or on-line in CG lab. If you use CG lab, try first 'tagger' and then 'morphological analyzer'. The latter will give your grammar ambiguous morphological input, while 'tagger' will provide disambiguated input.

```
(a) This is a test.

(b) When plants metabolize nitrogen compounds, they remove nitrogen from the water by
using it to build biomass. However, this is only temporary, as the plants release
nitrogen back into the water when older leaves die off and decompose. The most saline
water remains on the bottom, creating a barrier to the exchange of Oxygen and nutrients,
fostering totally different maritime environments. Watching animals of various sorts,
including a very wide range of birds, has also long been popular.
```

2. Basic-level annotation: Noun phrases

Next, try to make a little grammar for noun phrases, using the template grammar. The idea is to disambiguate part of speech by looking at the immediate np-context. Clear the grammar window, keeping only the DELIMITERS line and a SECTION or CONSTRAINTS header, and write:

```
SELECT (N) (-1 (ART)) ;
LIST PRE-N = ART DET ADJ STA ;
SELECT (N) (*-1 (ART) BARRIER (*) - PRE-N) ;
```

Run this on the following sentences:

```
I need a cold shower.
The biggest ship was built by Maersk.
```

To get morphological input for your sentences, use the CG lab or the relevant language page (English) on http://visl.sdu.dk (--> automatic analysis --> flat structure with menu settings for Parser=Analyzer pure, and Visualisation=Cohorts). Alternatively, you can get a sizeable wikipedia text cohort sample on the cg lab page (en link under "cohort sample"). Paste this in the IDE's input window, or save the file on your computer and open it in the IDE.

If you have a Linux system (or Mac), you can run files command line and evaluate the performance of your grammar:

cat wiki_test.cglab | vislcg3 -g yourgrammar --no-mappings > wiki_test.out

It is easy to quantify your disambiguation progress by counting cohort reading lines:

[r] *cat wiki_test_out | egrep -c ' "< ' (number of reading lines)*
[c] *cat wiki_test_out | egrep -c '^"<' (number of word forms)*
[p] *cat wiki_test_out | egrep -c ' PU ' (punctuation count)*

(r-p)/(c-p) is a measure for morphological ambiguity, and should ultimately approach 1 as you go on disambiguating. For later use, the CG lab page has also a gold standard version of the cohort sample, and a link to an evaluation program that will perform recall, precision and F-score evaluation.

3. Higher-level annotation: Place & Time semantic roles

Finally, using a full syntactic parser as input ('Standard Parser' in cg-lab, or "full syntactic parser" in http://beta.visl.sdu.dk/visl/en/parsing/automatic/parse.php), try to add semantic roles for place and time to prepositions (as place holders for the whole pp). Since at this level of analysis, parses already contain @tags (for syntax), you have to use a different prefix, the VISL convention being '§' or '%' for semantic roles. This can be achieved by by including the following line in the grammar:

```
MAPPING-PREFIX = §;
```

Try some of the following semantic roles:

§LOC = place/topological ... *in France*
§DIR = goal/direction .... *to France*

§ORI = origin, source ... *from France*

§LOC-TMP = point/period in time ... *in 1999, before 1999, between Tuesday and Thursday*
§DIR-TMP = time/temporal goal ... *until 1999, up=to 1999*
§ORI-TMP = time/temporal source ... *since 1999*
§EXT-TMP = temporal extension, durative/period ... *during the war, in 4 hours, for 10 minutes*

For a start, simply write MAP rules into the grammar window. These will be run *before* any disambiguation rules (REMOVE or SELECT) you might want to add later (in case you MAPping rules are ambiguous). For your time & place grammar, you can exploit the fact, that time/place prepositions should already be marked as @<ADVL, @ADVL>, @<AS, @<AO etc. by the syntactic parser. The following rule, for instance, will mark the preposition 'in' for place-hood, given a few set definitions:

```
LIST N/PROP-LOC = <top> <civ> <inst> <L> <Lh> <Lciv> <Lwater> <Lpath> <build> <BB>  ;
LIST @ADVL = @<ADVL @ADVL> @<SA @SA> @<OA @OA> ;

MAP (§LOC) TARGET N/PROP-LOC (0 @P<) (*-1 PRP LINK 0 (“in”) + @ADVL) ;
```

An example of a time rule:

```
MAP (§LOC-TMP) TARGET (NUM @P<) (*-1 PRP LINK 0 (“in”)) ;
```

The following sentences are from Wikipedia and the Leipzig corpus collection. All contain at least the preposition 'in', but you are of course welcome to mark other prepositions, too.

```
   Alexander Militarev suggests that their homeland was in the Levant.
   The settlement of Troy starts in the Neolithic, but continues up into the Iron age.
   Writings dealing with this subject are extant in Greek, Latin, Slavonic, Syriac, Armenian
and Arabic.
   It is also found, though in subordinate quantity, at Watchet in Somerset, near Penarth in
Glamorganshire, and elsewhere.
   Sect members won political control of the nearby community of Antelope, renaming it City
of Rajneesh, and attempted to control voting in Wasco County by busing thousands of homeless
people to the commune in 1984.
   From the names of their kings, it seems logical that the cult of the Baalim probably
coexisted in Ammon.
   They have a daughter, Rosie, and live in London.
   Such balloons, which lift a payload using buoyancy, should not be confused with balloons
in space, launched with a rocket, which are simply large deployable structures.
   Most vocal creationists are from the United States, and creationist views are much less
common elsewhere in the Western World.
   Many fled the country after the communists reunified it in 1975.
   In this approach, which sees copyleft primarily as a tool in a broadly scaled sniggling
operation, the intention of copyleft is to permanently minimize the restrictions imposed
under Intellectual Property regimes.
   Constantine married at some point in his life, but virtually nothing is known of it.
   Some stellar coronae, particularly in young stars, are much more luminous than the Sun's.
```

A rebel statement sent to Lisbon from Jamba said 86 government soldiers and 13 guerrillas were killed in the fighting that ended Jan. 3.

The East German news agency ADN reported Saturday that 70 segments of the Berlin Wall will be auctioned off on June 23 in the principality of Monaco.

They were printed while Mrs. Sutcliffe was pursuing a libel case against the magazine for alleging she tried to cash in on her husband's notoriety by agreeing to sell her story to a newspaper.

Polish living standards dropped sharply in the early 1980s and still have not climbed back to the level of a decade ago.

Lantos, 62, a member of the House since 1981, terminated an interview last October and requested that further questions be submitted in writing.

Jaruzelski's vote to lift the ban on Solidarity concluded his dramatic change of course in the past eight years.

Many residents believe the problem in Sindh is beyond the government's control.

USDA said sales of sorghum totaled 347,100 tons, the most in the current marketing year.

Viardo won the Van Cliburn International Piano Competition in Fort Worth, Texas, in 1973, and gave concerts in the United States in 1974 and 1975.

If he were brilliant, he probably would not be a family doctor in the first place.