

An Introduction to Grammar Engineering using HPSG: DAY 5

Background

Here are some possible translations of the English sentences from Day 4 $\frac{1}{2}$:

- (1) a. Ĉu ŝia onklino konis mian patrino?
'Did her aunt know my mother?'
- b. Lia saneco malboniĝis.
'His health has deteriorated.'
- c. La knaboj ankaŭ povas lerni malfacilajn lingvojn en la lernejo.
'The boys can also learn difficult languages at school.'
- d. La monaĥoj beligis la preĝejon.
The monks adorned the church.
- e. Ĉu via patrino volas dormigi la knabojn
'Does your mother want to put the boys to sleep?'

If these don't make sense to you, let one of us know. Before you start today's exercise, check out a small partial grammar of Esperanto:

`cvs checkout esperanto`

This grammar should look very familiar—it is very much like the English grammar you had completed yesterday. The lexicon has been replaced with an inventory of Esperanto roots, things have been set up for the more complex nominal morphology, and a couple of rules have been added. Also, to make typing easier, letters which have accents in standard Esperanto orthography are written as double letters for the purposes of this grammar. So, that means *ĉu* is written `ccu` and *ankaŭ* is written `ankauu`. Take a minute to get familiar with the grammar and figure out how it all works.

One thing to pay special attention to is different type of lexical rules. For English, we had a distinction between derivational rules which mapped lexemes to lexemes, and inflectional rules which mapped lexemes to words. This general plan will work for Esperanto verbs as well, since while they may have several derivational affixes (added by lexeme to lexeme rules), they only have one inflectional affix (added by a lexeme to word rule). For Esperanto nouns and adjectives, though, the situation is more complicated. A noun like *knabojn* 'boys (ACC)' consists of a root *knab* plus three inflectional endings *o*, *j*, and *n*. To handle this we need to add new kinds of lexical rules that we'll call lexeme to stem, stem to theme, and theme to word lexical rules. Every noun and adjective lexeme has to go through each of these layers before it can become a word:

<code>knab (lexeme) + o</code>	<code>=</code>	<code>knabo (stem)</code>
<code>knabo (stem) + j</code>	<code>=</code>	<code>knaboj (theme)</code>
<code>knaboj (theme) + n</code>	<code>=</code>	<code>knabojn (word)</code>

In addition, nouns and adjectives, like verbs, can undergo derivational (lexeme to lexeme) rules, adding even more complexity:

<code>knab (lexeme) + in</code>	<code>=</code>	<code>knabin (lexeme)</code>
<code>knabin (lexeme) + o</code>	<code>=</code>	<code>knabino (stem)</code>
<code>knabino (stem) + j</code>	<code>=</code>	<code>knabinoj (theme)</code>
<code>knabinoj (theme) + n</code>	<code>=</code>	<code>knabinojn (word)</code>

As in the English grammar, we control which rule can apply to which form using a set of lexical types. Play around with this a little to see how it works, but keep in mind that not all the lexical rules are in place yet. Some of them you'll be adding in the following exercises.

Exercises

1. One linguistic device that plays an important role in Esperanto but almost none in English is case marking. Nouns which function as a direct object take the accusative suffix *-n*, and adjectives and possessive determiners agree in both case and number with the noun they modify.
 - (a) Add a feature *CASE* to *nominal* which takes a feature structure of type *case* as its value. Add *nom* and *acc* as subtypes of *case*.
 - (b) Add constraints to the relevant lexeme types to enforce case assignment. Verbs select for a nominative specifier and (if transitive) an accusative complement, and prepositions select for a nominative complement.
 - (c) Add case agreement constraints so that determiners and adjectives agree with the noun in case as well as number.
 - (d) Add the lexical rules for case marking. Since the case marker comes last in the word, case marking rules are theme to word lexical rules. The grammar you checked out has a lexical rule *no-case-word-rule* in `inflr.tdl` and a corresponding type *no-case-word* in `types.tdl`. This rule maps any theme to a word without adding a case ending. Delete this rule (it's just a placeholder so you can load the grammar), and add the correct lexical rules which map a theme to either a nominative word or an accusative word, with the appropriate ending. You will probably need two rules, one for each case.
 - (e) Check your work by parsing the sentences in `case.items`.
2. Next you need to add the possessive determiners. Esperanto possessives are tricky because syntactically they behave like determiners, but morphologically they behave like adjectives.
 - (a) Create a subtype of *det* called *poss*. Since the inflectional rules for adjectives will have to apply to either adjectives or possessives, add *adj-or-poss* as a subtype of *nominal* and make both *adj* and *poss* inherit from it.
 - (b) Modify *adj-stem* so that the *adj-stem-rule* will apply to both adjectives and possessives.
 - (c) Add a lexeme type for possessive determiners.
 - (d) Add new lexical entries for the possessive forms of the pronouns in `lexicon.tdl`. (Or, if you are feeling ambitious, set up a lexical rule to derive possessives from pronouns, or to derive pronouns from possessives.)
 - (e) Check that your modifications have had the desired effect by parsing the examples in `poss.items`.
3. Esperanto also has a number of lexeme to lexeme lexical rules. Some, like *mal-* prefix rule, map a lexeme to a lexeme of the same category, but with a different meaning. Others, like the *-igg* suffix rule, map a lexeme of one category to a lexeme of a different category. Following the example of the *neg-adj-rule* and the *inchoative-verb-rule*, add lexical rules for the suffixes *-in*, *-ej*, *-ec*, and *-ig*. You may find it easier to add two separate rules for *-ig*, one mapping adjectives to verbs and one mapping verbs to verbs. Test your results with `endings.items`.

4. Infinitives (like *lerni* ‘to learn’) are verb forms, but infinitive phrases (like *lerni la esperantan lingvon* ‘to learn the Esperanto language’) are a lot like noun phrases in that they can be the subject or object of a verb. Add a lexeme to word lexical rule for infinitives which forms a noun word from a verb lexeme. The HEAD value of the output should be *noun*, but the COMPS value should be the same as the underlying verb and the *spr* value should be empty. Test your results with `inf.items`.
5. Add new lexeme types and lexical entries for the particles *ĉu*, *ne*, and *ankaŭ*. These particles should combine with the phrases they attach to via the head-complement rule. Make a set of test items to verify your analysis.
6. Finally, try parsing the sentences from last night’s exercise in `hard.items`. Does everything work? If not, try to figure out what’s going wrong. If so, then great: you’ve now got a fairly wide coverage syntactic grammar of Esperanto. All that’s left to do is add semantics. (If you have time today, see how far you can get with it. Otherwise, that’s maybe a project to come back to on some snowy February afternoon.)