Computational Linguistics (Spring 2011) — Exercise 3b

1 A New Language: Esperanto (20 Points)

Esperanto is an artificial language created by the Polish oculist L. L. Zamenhof in the nineteenth century. Zamenhof designed the language to have a very simple and regular morphology, on the theory that war and pestilence were caused by irregular verbs. While his dream of world peace has not been achieved, his language has turned out to be very useful for constructing problem sets in introductory linguistics classes.[1]

For this exercise, you will be applying the grammar engineering techniques you have mastered for English to a small grammar of Esperanto. To get ready for that, try out this pencil-and-paper exercise first. Based on the sentences in (1), write a small English–Esperanto morpheme 'dictionary' (a table of translations at the level of morphemes), and use it to translate the English sentences in (2) into Esperanto. Working through this will give you a feel for the language and you will find building the actual Esperanto grammar easier to follow. Please hand in your morpheme dictionary (preferably in electronic form) as part of your submission.

(1) a. La alta knabo malsaniĝis.
         'The tall boy fell ill.'

     b. Ĉu li grandigis la grandecon de la dormejo?
         'Did he increase the size of the dormitory?'

     c. Ankaŭ malaltaj knabinoj povas esti belaj.
         'Short girls, too, can be beautiful.'

     d. Mia patro estas sana ĉar li ne trinkas vinon.
         'My father is healthy because he doesnt drink wine.'

     e. La bonaj monaĥinoj volis preĝi en la preĝejo.
         'The good nuns wanted to pray in church.'

     f. Lerni la esparantan lingvon estas facila.
         'It's easy to learn Esperanto.'

     g. Mi vidis ŝian onklon en la trinkejo.
         'I saw her uncle in the bar.'

     h. La beleco de la lingvo estas ĝia facileco.
         'The beauty of the language is its simplicity.'

     i. Ĉu vi konas miajn onklojn?
         'Do you know my uncles?'

(2) a. Did her aunt know my mother?

     b. His health has deteriorated.

     c. The boys can also learn difficult languages at school.

     d. The monks adorned the church.

     e. Does your mother want to put the boys to sleep?

The first step in analyzing a new language like this is to figure out which words in Esperanto match up to which words in the English translations. Look for sentences with the same or related words in the English translation, and then look for words which look similar in the corresponding Esperanto sentences. In this case, it's pretty easy, since Esperanto's syntax is almost exactly the same as that

---

[1]Warmest thanks to Rob Malouf, professor at San Diego State University, for his profound knowledge of its grammar and for creating this exercise.

of English (Zamenhof didn't have a lot of imagination). To take the first sentence as an example, we can break it down like this:

La	alta	knabo	malsaniĝis.
The	tall	boy	fell ill

Once you have a good idea what the rules for constructing sentences are, look for patterns across individual words. Find words with similar meanings or similar grammatical functions, and see if there are common elements in their spellings. Keep in mind that Esperanto has no irregular morphology, so stems and endings always have exactly one spelling. Working from the first example again, you might notice that adjectives often end in *a* and nouns often end in *o*, so these would be good candidates for adjective and noun endings. You might also notice a similarity between *malsaniĝis* 'fell ill' and *malaltaj* 'short' in example (1c), *sana* 'healthy' in (1d), and *volis* 'wanted' in example (1e). From this you might guess that the structure of this word is:

mal	san	iĝ	is
not	healthy	become	past

= became not healthy

Continue in the same way with all the words, and you will collect an inventory of morphemes that you can use to construct the new words you need to translate the sentences in (2).

2 Part of the Solution: English Translations (0 Points)

Here are some possible translations of the English sentences for the Esperanto examples:

(3) a. Ĉu ŝia onklino konis mian patrinon?
     '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 after a bit of study, please go back to the Esperanto data and find help, if necessary; there are many on-line resources on Esperanto on the Internet.

3 Obtaining the Starting Grammar (0 Points)

To begin our final exercise, once logged into the IFI Linux environmnet, obtain a small partial grammar of Esperanto. At the shell prompt, execute the following command:

```
esperanto
```

This grammar, located in the sub-directory 'esperanto', should look familiar—it is very much like the English grammar you have been working with. The lexicon has been replaced with an inventory of Esperanto stems, 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 $\hat{c}u$ is written `ccu` and *ankaŭ* is written `ankauu`. Take a few minutes to get familiar with the grammar and figure out how it works.

One thing to pay special attention to is different types 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 stem *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:

$$\begin{array}{lll}
\text{knab (lexeme)} + \text{o} & = & \text{knabo (stem)} \\
\text{knabo (stem)} + \text{j} & = & \text{knaboj (theme)} \\
\text{knaboj (theme)} + \text{n} & = & \text{knabojn (word)}
\end{array}$$

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

$$\begin{array}{lll}
\text{knab (lexeme)} + \text{in} & = & \text{knabin (lexeme)} \\
\text{knabin (lexeme)} + \text{o} & = & \text{knabino (stem)} \\
\text{knabino (stem)} + \text{j} & = & \text{knabinoj (theme)} \\
\text{knabinoj (theme)} + \text{n} & = & \text{knabinojn (word)}
\end{array}$$

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.

4 Case Marking (30 Points)

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.

1. 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*.

2. 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.

3. Add case agreement constraints so that determiners and adjectives agree with the noun in case as well as number.

4. 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 'irules.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.

5. Check your work by parsing the sentences in the file 'case.items'.

5 Possessive Determiners (20 Points)

Next you need to add the possessive determiners. Esperanto possessives are tricky because syntactically they behave like determiners, but morphologically they behave like adjectives.

1. 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.

2. Modify *adj-stem* so that the *adj-stem-rule* will apply to both adjectives and possessives.

3. Add a lexeme type for possessive determiners.

4. 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.

5. Check that your modifications have had the desired effect by parsing the examples in the file 'poss.items'.

6 (More) Derivational Rules (50 Points)

Esperanto also has a number of lexeme-to-lexeme lexical rules. Some, like the *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.

1. 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 easer to add two separate rules for *-ig*, one mapping adjectives to verbs and one mapping verbs to verbs. Test your results with the file 'endings.items'.

2. Infinitives (like *lerni* 'to learn') are verb forms, but infinitive phrases (like *lerni la esparantan 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 the file 'inf.items'.

3. 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.

4. Finally, try parsing the sentences from the first part of this exercise, which you'll find in the file '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 broad-coverage syntactic grammar of Esperanto!

7 Submitting Your Results

To provide your results to us, please pack up the entire contents of your 'esperanto/' directory when you are done. Email the archive file to both Lars-Erik and Stephan before the final deadline, midnight on Friday, May 20. In the IFI Linux environment, we provide the *submit* command-line tool for you to automate the process of packaging up and sending your results to us.