ESSLLI'99, Utrecht, August 9–20, 1999
M. Butt, A. Frank, J. Kuhn:
Development of large scale LFG grammars:
Linguistics, Engineering and Resources

French LFG: Exercises 3

– Tokenization, morphology, and sublexical rules –

Copy the grammar
`/space/projects/esslli99/grammars/french-sublex/french-sublex.lfg`.

Exercise 1

The grammar is now interfaced with tokenization and morphology for French. The relevant finite-state transducers are referenced in the grammar configuration.

For some categories you already find sublexical rules and lexical entries in the grammar sections `FRENCH SUBLEX RULES` and `FRENCH SUBLEX LEXICON`.

Lexical entries for nouns, verbs, prepositions and (new!) adverbs are now only specified for stem forms. Try to understand the differences that arise between the previous grammar version without sublexical rules and the new version, by looking at template definitions and e.g. the definition of agreement constraints.

Test how the grammar behaves now. Use "unknown" vocabulary for the categories that are described by generic Lunknown lexical entries. You may also introduce further generic Lunknown lexical entries.

- Why are there no Lunknown entries for verbs?
- Proper names are now analyzed as `N`. What has been changed in the `NP` rule?
- Why do we now get 2 analyses for *Marie aime le chat* and *La fille aime le chat*? Restrict the grammar.
- Observe the effect of tokenization by parsing examples like `PP: aux enfants`, `PP: des enfants`. (For `NP: des enfants` (e.g. in *Jean voit des enfants*) we will have to introduce indefinite determiners (see below).)
- The tokenizer optionally decapitalizes the first word of a sentence. You can now type in sentences as `Le chat dort`. However, decapitalization causes an unwarranted ambiguity for *Jean dort*. Lunknown lexical entries can be declared *suboptimal* in terms of so-called optimality marks. We will discuss

optimality-style constraint ranking in detail only later. Aor the time being, we will filter the unwarranted ambiguities by the following additions:

Please add:
– to the CONFIG section: OPTIMALITYRANKING: NEUTRAL Unknown NOGOOD.
– to the LUnknown lexical entry for N: Unknown $ o::*

- *Le chat dort.* is assigned MOOD = INDICATIVE, whereas *Le chat aime Marie.* is not. Inspect the sublexical nodes in both cases, and discuss the consequences of an alternative treatment, where mood is specified in terms of disjunction instead of being underspecified.

- Why do we obtain 2 analyses for *Jean voit la fille qui aime le chat*? Restrict the grammar.

Aufgabe 2: Sublexical rules

Define the morphology interface for clitics and determiners, following the model of the existing sublexical rules. The lexical entries will have to be marked by XLE instead of *.
The relevant morphological analyses for clitics and determiners are stated below. You can obtain these morphological analyses in XLE with the command
analyze-string <string>.

```
je   je+Nom+InvGen+SG+P1+PC
tu   tu+Nom+InvGen+SG+P2+PC
il   il+Nom+Masc+SG+P3+PC
elle il+Nom+Fem+SG+P3+PC
ils  il+Nom+Masc+PL+P3+PC

le   le+Masc+SG+Def+Det
     le+Acc+Masc+SG+P3+PC
la   le+Acc+Fem+SG+P3+PC
     le+Fem+SG+Def+Det
les  le+InvGen+PL+Def+Det
     le+Acc+InvGen+PL+P3+PC

lui  lui+Dat+InvGen+SG+P3+PC
leur leur+Dat+InvGen+PL+P3+PC

un  un+Masc+SG+Indef+Det
une un+Fem+SG+Indef+Det
des un+InvGen+PL+Indef+Det
```