Homework 1

Concepts and Applications in NLP

October 31, 2024

1 Assignment 1

Esperanto, created in 1887, is the world's most widely spoken constructed international auxiliary language. It is based on roots from various European languages and retains the structure of these languages. One of the language's most notable features is its extensive system of derivation, where prefixes and suffixes may be freely combined with roots to generate words, making it possible to communicate effectively with a smaller set of words.[1]

Some Esperanto facts:[2]

`hund`	*dog*
`kat`	*cat*
`bird`	*bird*
`elefant`	*elephant*

1. Esperanto nouns begin with a root
2. A bare root, like *hund* is not valid
3. All nouns must have a suffix, like *hundo, elefanto*
4. Gender:
   - *elefanto*: masculine or unmarked for gender
   - mark as feminine by placing the suffix *in* between root and *o* suffix: *elefantino*
5. Diminutive/Augmentative:
   - mark as diminutive by placing the suffix *et* between root and *o* suffix:
     elefanteto: little elephant
   - mark as augmentative by placing the suffix *eg* between root and *o* suffix:
     elefantego: big elephant
6. Augmentative, diminutive and feminine suffixes can co-occur; it is not clear if there are any rules to limit their mutual order or co-occurrence.
   For now: use a loop to allow any number of these three suffixes in any order between the root and the o suffix.
7. Plural: suffix *j*, which appears directly after the *o* suffix (*birdoj: birds*)
8. Accusative:
   - suffix *n* after the *j* suffix (if present) or after the *o* suffix
   - no other morphemes can occur after the accusative *n*

[1] cf. `https://en.wikipedia.org/wiki/Esperanto`
[2] From Beesley and Karttunen (2003)

Write a `lexc` grammar to model Esperanto nouns.

Start your grammar like this:

```
LEXICON Root
Nouns ;

LEXICON Nouns
bird Nmf ;
hund Nmf ;
kat Nmf ;
elefant Nmf ;
```

Hint: for (6), you can refer to the current `LEXICON` to create a loop.
Remember to provide a path to another `LEXICON` to end the loop!

Test your grammar. It should only accept or reject:

- **accept** birdo, birdoj, birdon, birdeton, birdetinegojn, elefantegegegoj, hundon, hundojn, hundetojn
- **reject** hund, hundjo, katoego, elefantonj

Submit your grammar as `esperanto-yourname.lexc`

2 Assignment 2

Define a lexicon and a set of rules to construct a transducer to model a fragment of an artificial language with the following properties:

1. Nouns consist of a stem and a suffix for either feminine (*la*) or masculine *to*.
2. Nouns can have an optional suffix for plural: *fi*.
3. Verbs can have an optional negation prefix *mon*
4. Enclosing a verb stem by the morpheme *na* transforms the verb into an adjective
5. Such a derived adjective can have an additional suffix for comparative (*be*) or superlative (*ke*)
6. Verb tenses are represented by a suffix:
   - Present tense: no further suffix
   - Past: *hi*
   - Future: *il*
7. When adding the past tense morpheme *hi*, verbs ending with a consonant need to insert an *e* before *hi*
8. When adding the future morpheme *il*, verbs ending with a vowel need to insert an *l* befor *il*

Vocabulary:

verb stems	**noun stems**
rag	topo
veg	gran
kise	isna
lige	peco

Define a `lexc`-grammar to model (1) - (6).
You can start your grammar like this:

```
Multichar_Symbols +Noun+Masc +Noun+Fem +Verb +Pres +Past +Fut
+Sg +Pl Neg+ +Adj+Comp +Adj+Sup

Definitions
VerbStem = [ {rag} | {veg} | {kise} | {lige} ] ;

LEXICON Root
Nouns ;
Verbs ;
VerbPref ;
```

Hint: for (4) you can use an entry of the form

```
< 0:{na} VerbStem 0:{na} >        Adj;
```

where `Adj` is the continuation class for a new `LEXICON`.

Implement (7) and (8) as rules in a `foma` script.

Test your grammar:

rag+Verb+Fut	ragil
rag+Verb+Past	ragehi
rag+Verb+Pres	rag
rag+Adj+Sup	naragnake
rag+Adj+Comp	naragnabe
rag+Adj	naragna
Neg+lige+Verb+Past	monligehi
Neg+lige+Verb+Pres	monlige
Neg+lige+Verb+Fut	monligelil
Neg+kise+Adj+Sup	monnakisenake
Neg+kise+Adj+Comp	monnakisenabe
Neg+kise+Adj	monnakisena
topo+Noun+Fem+Pl	topolafi
topo+Noun+Fem+Sg	topola
topo+Noun+Masc+Pl	topotofi
topo+Noun+Masc+Sg	topoto
???	topo
???	kiseil
???	raghi

Submit your `lexc`-file and your `foma`-file.

Please submit your homework by mail on or before **14. November 2024**
to `marion.dimarco -AT- tum.de`