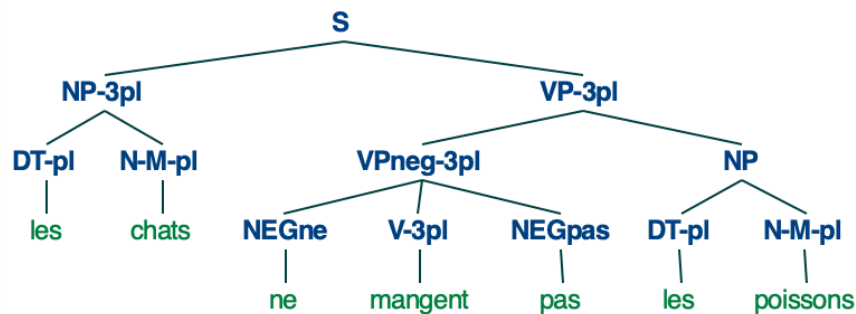


COMP 550 Assignment 2  
Daniel Korsunsky, ID: 260836959  
November 15, 2021

Example sentences accepted by CFG and CYK parser:

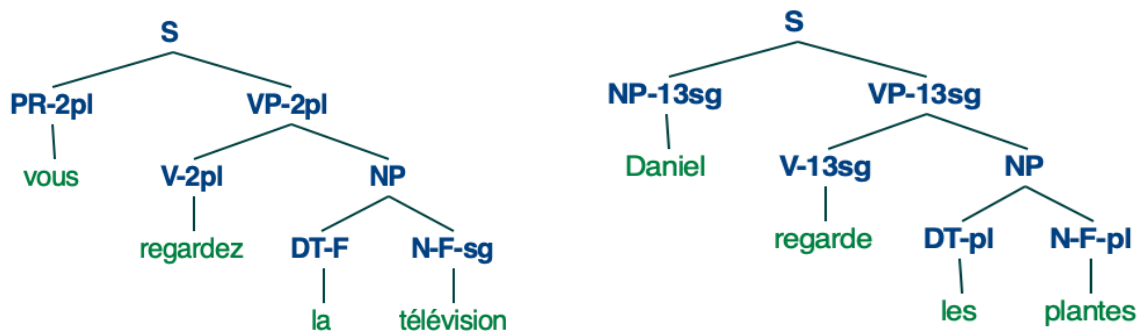
(1) Les chats ne mangent pas les poissons. (The cats do not eat the fish (Pl).)

```
(S
  (NP-3pl (DT-pl les) (N-M-pl chats))
  (VP-3pl
    (VPneg-3pl (NEGne ne) (V-3pl mangent) (NEGpas pas))
    (NP (DT-pl les) (N-M-pl poissons))))
```



(2) Vous regardez la télévision. (You (formal) watch the television.)

```
(S
  (PR-2pl vous)
  (VP-2pl (V-2pl regardez) (NP (DT-F la) (N-F-sg télévision))))
```



(3) Daniel regarde les plants. (Daniel watches the plants)

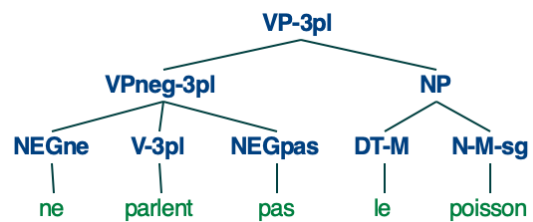
```
(S
  (NP-13sg Daniel)
  (VP-13sg (V-13sg regarde) (NP (DT-pl les) (N-F-pl plantes))))
```

### Example sentences rejected by CFG and CYK parser (returned by .parse() method):

- ERROR: No valid syntax for "je mangent le poisson" in given CFG
- ERROR: No valid syntax for "je mange les" in given CFG
- ERROR: No valid syntax for "la poisson mangent les chats" in given CFG

### Example of overgeneration:

A case of overgeneration is one in which the CFG generates a sentence as grammatical when it is in fact not. My CFG can generate phrases such as “ne parlent pas le poisson” (“do not speak fish”) which can be seen as syntactically correct but (i) is not a complete sentence, and (ii) is semantically nonsensical.



### Example of undergeneration:

A case of undergeneration is one in which the CFG does not generate a sentence that is in fact grammatical. Since the CFG does not handle prepositions, it naturally rejected the sentence “Je suis au Canada” (“I am in Canada”).

### What are some advantages of modeling French grammar with a CFG?

Especially for a non-native speaker like myself, modeling a foreign language with a CFG was useful because it is easy to understand, write, and parse sentences with, especially once converted to CNF. Accordingly, it was easy to create subcategories for various nonterminals to account for aspects like verb and plurality agreement. This ensured that the CYK could successfully parse the sentence without having to engineer complex grammar rules.

### What are some disadvantages of modeling French grammar with a CFG?

For one, CFGs do not take into account attachment height, as attachments follow the same production rules regardless of height. Moreover, CFGs cannot really restrict production rules to specific contexts, so I was forced to produce some very specific nonterminal categories (for example, plural masculine adjectives that appear after nouns) in order to account for the complexity in French syntactic agreement.

### What are some aspects of French grammar that your CFG does not handle?

My CFG cannot handle any instances of prepositions, prepositional phrases, conjunctions, adverbs, or temporal markers, to name a few. The only verbs it can account for are in the present tense, which excludes imparfait and passé composé, for instance.

The grammar (and parser) are also unable to handle French contractions such as *j'aime* (instead of *\*je aime*) that occur when a word starts with a vowel.

*Note regarding my `to_CNF()` method in the CYK class:*

The examples generated in this document were converted to CNF using the built-in `nltk.CFG` method, but converted back from CNF in the `parse()` function I wrote. As you will see in my code, you can toggle between the built-in method and the one I had written in the CYK class. I had gotten my own method working on an imported English language CFG with all three CNF rules, and believe that the logic is mostly solid. However, once I started to test it on the CFG I came up with for Q1, I started running into some errors and was unable to resolve them in time. I hope that my own method can still be recognized for my attempt. Thank you!