

# ASSIGNMENT 2

COMP 550, Fall 2021

Due: Monday, November 15<sup>th</sup>, 2021, 9:00pm.

You must do this assignment individually. You may consult with other students orally, but may not take notes or share code, and you must complete the final submission on your own.

Question 1: 60 points

Question 2: 40 points

---

100 points total

## Assignment

### Question 1: Grammar for French (60 points)

In this question, you will develop a context-free grammar for a fragment of French. Your grammar must account for various aspects of the French language, as listed below.

#### Basic sentence word order in the present

The basic word order in French is Subject-Verb-Object, as in English:

- (1) Je regarde la télévision.  
I watch the television
- (2) Le chat mange le poisson.  
The cat eats the fish

#### Subject-verb agreement

Just as in English, the subject must agree with the verb in number and person:

- (3) Tu regardes la télévision.  
You(2Sg) watch the television
- (4) Il regarde la télévision.  
He watches the television
- (5) Nous regardons la télévision.  
We watch the television
- (6) Vous regardez la télévision.  
You(2Pl) watch the television
- (7) Ils regardent la télévision.  
They(Masc.) watch the television

Look up the list of subject pronouns in French, as well as the verb conjugation paradigm for several common verbs using an online website. Include these in your grammar.

**Reference:** <http://www.wordreference.com/conj/FrVerbs.aspx>

## Negation

Basic verb negation is handled in French by surrounding the verb with the particles *ne ... pas*.

- (8) Tu ne regardes pas la télévision.  
You(2Sg) do not watch the television
- (9) Le chat ne mange pas le poisson.  
The cat does not eat the fish

## Definite noun phrases and proper names

A definite noun phrase in French follows a similar order as in English (article + noun). However, the article must agree with the noun in number and grammatical gender. Grammatical gender is a more-or-less arbitrary categorization of nouns into either masculine or feminine.

Examples:

- (10) Le chat  
the(Masc.) cat
- (11) La télévision  
the(Fem.) television
- (12) Les chats  
the(Pl.) cats
- (13) Les télévisions  
the(Pl) televisions

As you can see, there is no distinction in the plural between masculine or feminine.

Some proper names in French do not take articles, just as in English:

- (14) Jonathan  
Jonathan
- (15) Montréal  
Montreal

Others do (e.g., *le Canada*), which you have to handle:

- (16) le Canada  
Canada

You can look up examples of proper nouns which need an article. e.g. <https://language-easy.org/french/grammar/nouns/proper-nouns/>

## Direct object pronouns

When a pronoun is a direct object of the verb, they precede the verb:

- (17) Il la regarde.  
He it(Fem.) watches.

Look up the list of direct object pronouns in French, and enhance your grammar to account for the word order with direct objects. <https://grammar.collinsdictionary.com/french-easy-learning/personal-pronouns-direct-object>

## Attributive adjectives

Adjectives typically follow the noun that they modify in a noun phrase:

(18) Le chat noir  
the(Masc.) cat black

(19) Le chat heureux  
the(Masc.) cat happy

However, other adjectives precede the noun:

(20) Le beau chat  
the(Masc.) beautiful cat

(21) Le joli chat  
the(Masc.) pretty cat

Yet others may precede OR follow the noun, though the meaning usually changes slightly:

(22) La dernière semaine  
the(Fem.) last week  
*the last week (e.g., of the year)*

(23) La semaine dernière  
the(Fem.) week last  
*last week (i.e., the one before this week)*

In addition, adjectives must agree with the noun that they modify in number and gender:

(24) Les chats noirs  
the(Pl.) cats black(Pl.)  
*the black cats*

(25) La télévision noire  
the(Fem.) television black(Fem.)  
*the black television*

(26) Les télévisions noires  
the(Pl.) televisions black(Fem. Pl.)  
*the black televisions*

Note that adjectives do distinguish masculine from feminine in the plural.

Find several adjectives of each of the three classes above, and incorporate them into your grammar.

## References

<http://french.about.com/od/grammar/a/adjectives.htm>

[http://french.about.com/od/grammar/a/adjectives\\_4.htm](http://french.about.com/od/grammar/a/adjectives_4.htm)

## Examples and submission format

You already have many examples that your grammar should accept (though many of the above examples were only noun phrases, not full sentences). Here are some sentences that your grammar should **reject**:

(27) \*Je mangent le poisson.

(28) \*Les noirs chats mangent le poisson.

(29) \*La poisson mangent les chats.

(30) \*Je mange les.

Use the following nonterminals to indicate grammatical categories:

S	sentence/clause
NP	noun phrase

VP	verb phrase
N	noun
PN	proper noun
PR	pronoun
V	verb
DT	determiner
A	adjective

You may add further non-terminal categories or subdivide them (e.g., V-1Sing) as needed. Don't forget the lexical rules! Include enough lexical items such that each of the syntactic categories can be expressed in at least three different ways.

Write your grammar in a text editor using a predictable, computer-readable format. In particular, you must use a format that is compatible with NLTK's `CFG.fromstring` function. For instance, here is one possible rule:

`S -> NP VP`

Here is another example of a set of four rules (here, they are lexical rules):

`V-1Sg -> 'mange' | 'aime' | 'regarde' | 'cherche'`

These are just examples, and are not necessarily the rules you want in your grammar! Ignore punctuation and capitalization in your grammar (just use all lower-case, except for proper names). French has contractions in many cases where a word begins with a vowel (e.g., *j'aime* rather than *\*je aime*). You may ignore such issues.

Submit your grammar as a plaintext `.txt` file. Show instances where your grammar correctly accepts and rejects some sentence. In addition, show at least one case of overgeneration and one case of undergeneration in your grammar. (Hint: if you don't speak French, ask Google Translate or a friend for help! Also, many French grammar sites will give examples of mistakes.) In addition, answer the following questions in your response to the question:

1. What are some advantages of modelling French grammar with a CFG?
2. What are some disadvantages of modelling French grammar with a CFG?
3. What are some aspects of French grammar that your CFG does not handle?

This question is rather open-ended; your grammar will be judged on the following points:

- Whether you followed the specifications above (e.g. names of non-terminals, minimum number of lexical entries)
- Coverage of the required grammatical constructions
- Clarity of the grammar
- The responses to the questions above

You won't get extra points for having many additional lexical items that exhibit the same type of behaviour!

## Question 2: Implement CYK for your grammar (40 points)

Implement your own version of the CYK algorithm. Your code should define a class which represents a CYK parser. Your parser's constructor must take in a CFG grammar that is an instance of NLTK's `CFG` class. It must include a `.parse()` method that is able to take in sentences as a string with tokens separated by whitespace and run the CYK algorithm on it. The output of the method should be a list of all the possible trees of the sentence, which are instances of `nlk.tree.Tree`. This list may be empty if the sentence is not accepted by the CFG.

Run your parser on several sentences from your grammar from Q1, showing sentences that are accepted and sentences that are rejected by your grammar. Include these in the written portion of your submission.

Notes:

- You will also have to write code that converts your grammar to/from Chomsky Normal Form yourself.
- Be sure the final output parse is not in the CNF-version of the CFG!
- You can use any other parser or method from NLTK to check your work, but you must write your own implementation of the CYK algorithm.
- You may want to do Question 2 first before Question 1, as having a parser will be useful for checking your work!

## What To Submit

Your submission must be made through MyCourses, and must consist of the following three files:

1. The grammar file of Question 1, as a plaintext .txt file called 'french-grammar.txt'.
2. Your code for Question 2, 'a2-cyk.py'.
3. The written portions for both questions as a .pdf file called 'a2-written.pdf'.