

ASSIGNMENT 2

COMP 550, Fall 2018

Due: Tuesday, October 16th, 2018, 11:59pm.

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: 30 points

Question 2: 40 points

Question 3: 30 points

100 points total

Assignment

Question 1: POS Tagging (30 points)

Train a first-order (i.e., the probability of a tag depends only on the previous tag) HMM part-of-speech tagger by hand on the following training corpus. Find the MAP estimate of the parameters of the model using add-1 smoothing.

That/C that/N is/V , is/V .

That/C that/N is/V not/N , is/V not/N .

Is/V that/N it/N ?

Is/V it/N that/N good/J ?

Ignore capitalization differences and punctuation. They are only there for readability purposes. There should thus be a tagset of size 4, and a lexicon of size 5. Your model should contain the following parameters: Π (initial state probabilities), A (state transition probabilities), and B (emission probabilities).

Next, run the Viterbi algorithm (by hand) to obtain a POS tagging for the following test sentence. Show the trellis cell values and your calculations.

That is good .

Suppose that we now receive two additional unlabelled sentences. Note that there is now an additional lexical item, “bad”.

Bad is not good .

Is it bad ?

Update your previously trained model with a semi-supervised method involving Expectation Maximization over the additional sentences such that it can now account for the word “bad”. In particular show how you need to modify the previous model, then continue training with the new sentences. Show your calculations for one iteration of EM. Your solution should assign non-zero probability to unseen sentences involving the word “bad” and the other 5 lexical items, such as the sentence “That is bad .”

Question 2: Grammar for French (40 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>

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:

- (8) Le chat
the(Masc.) cat
- (9) La télévision
the(Fem.) television
- (10) Les chats
the(Pl.) cats
- (11) 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:

- (12) Jackie
Jackie

- (13) Montréal
Montreal

Others do (e.g., *le Canada*), but you do not have to handle them.

Direct object pronouns

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

- (14) 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.

Attributive adjectives

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

- (15) Le chat noir
the(Masc.) cat black

- (16) Le chat heureux
the(Masc.) cat happy

However, other adjectives precede the noun:

- (17) Le beau chat
the(Masc.) beautiful cat

- (18) Le joli chat
the(Masc.) pretty cat

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

- (19) La dernière semaine
the(Fem.) last week
the last week (e.g., of the year)

- (20) 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:

- (21) Les chats noirs
the(Pl.) cats black(Pl.)
the black cats

- (22) La télévision noire
the(Fem.) television black(Fem.)
the black television

- (23) 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

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**:

(24) *Je mangent le poisson.

(25) *Les noirs chats mangent le poisson.

(26) *La poisson mangent les chats.

(27) *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. 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, answer the following questions in your response to the question:

1. What are some advantages of modelling French grammar with a CFG, compared to using an FSA?
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 3: Reading Assignment (30 points)

Read the following paper:

Dan Klein and Christopher D. Manning. Accurate Unlexicalized Parsing *ACL 2003*. <http://www.aclweb.org/anthology/P03-1054>

Write a max. one-page (c. 500 words) discussion on this paper, including the following points:

1. A brief summary of the contents of the paper, including the theoretical framework and the experiments. Define what lexicalized and unlexicalized parsing mean in this paper.
2. An evaluation and synthesis of what you learned in the paper. What are the advantages and limitations of this work? How does it relate to the topics that we have discussed in class?
3. Three questions related to the paper. These can be clarification questions, or questions about potential extensions of the paper, or its relationship to other work.