University of Toronto, Department of Computer Science
CSC 2501/485F—Computational Linguistics, Fall 2015

Assignment 1

Due date: 13:10, Monday 5 October 2015, on CDF.
This assignment is worth 12% of your final grade.

- Submit the ID file on the course website along with your submission.
- Please type your answers in no less than 12pt font; diagrams and tree structures may be drawn with software or neatly by hand (or by other extremity).
- Any clarifications to the problems will be posted on the course website announcements section. You are responsible for taking into account in your solutions any information that is posted there, or discussed in class, so you should check the page regularly between now and the due date.
- What you turn in must be your own work. You may not work with anyone else on any of the problems in this assignment, except for discussion in very general terms. If you need assistance, contact the instructor or TA.

1. What is "understanding"? [10 marks]

Think about what is involved in understanding each of the following texts (to the extent that each is understandable). *Briefly* explain the potential difficulties (if any) in each of the following texts for a computer "understanding" process. Which texts (if any) might be easier for a computer than for a human?

a. Two penguins are on an ice-floe. One penguin says to the other, "You look like you're wearing a tuxedo." And the second penguin says, "How do you know I'm not?"

b. How do I love thee? Let me count the ways.
I love thee to the depth and breadth and height
My soul can reach, when feeling out of sight
For the ends of Being and ideal Grace.
—*Elizabeth Barrett Browning* (1806–1861)

c. For a partnership, the employer is considered to be a corporation where all the shares with full voting rights are owned by the members in the same proportion as the member's share of the income or loss of the partnership. For an individual, the employer is considered to be a corporation where the individual owns all the shares with full voting rights of the corporation's capital stock. The allocation of the premium relief between associated employers is determined in the following way: When a written agreement exists, the allocation of the premium relief is determined by the terms of this agreement, if it is signed by all the associated employers and filed with the Minister of Revenue.

d. Sunday, August 12

nt really much happen,

the normal routine, had breakfast
then, make the npcc stuff,
think i m gettin crazy over polishin boots now ..

for the whole of afternoon,
pratice maths ..
i juz walk to the kitchen to have a drink ..
or juz go n take some thing n my mum is like askin for trouble,
so kind of fought wif her ..
unreasonable, controllin myself to cool down ..
walao, she is really unstandable ..

later carried on wif my maths ..
spoil my mood of doin work ..
do till evening 6plus ..

tidy up my room .. kind of a bit untidy ..

aft dinner,
slack in front of the com ..
sian aft a long holiday need to return to sch,
afraid tat would b able to wake up ..
the most spooky thing is
the gate is open at 12midnight .. x((

e. Ross wants to marry a millionaire.

f. *Today:* Increasing cloudiness. 40 percent chance of showers this afternoon with the risk of a thunderstorm. Wind becoming west 30 km/h this afternoon. High 30. UV index 7 or high. *Tonight:* Cloudy with 40 percent chance of showers early this evening. Risk of a thunderstorm. Clearing this evening. Wind west 20 km/h. Low 18. *Monday:* Sunny. Wind northwest 30 km/h. High 26.

2. Grammars [10 marks]

The grammar given below covers only **declarative** sentences: those that assert a proposition:

> Nadia walks her dog in the park.

Two other kinds of sentences in English are *imperative* (commands):

> Walk your dog.

and *interrogative* (questions):

> Who walks her dog in the park?
> What does Nadia walk in the park?
> Where does Nadia walk her dog?
> Does Nadia walk her dog in the park?

Modify the grammar by adding or changing rules so that it accounts for these kinds of sentences. Your grammar should be reasonably principled and produce good sentences without producing bad ones. For example, it should **not** produce sentences such as these:

> (??) What Nadia walk in the park?
> (??) What does Nadia walk her dog in the park?
> (??) Where walks her dog in the park?

(where the '(??)' indicates ill-formedness).

Note: You do **not** need to account for verb endings due to person and tense (such as the difference between *walk* and *walks* or *did* and *does*). So your grammar might produce sentences such as *Does I walks my dog*. That's okay; we'll see how to deal with this problem later. But your grammar should **not** produce very non-English sentences such as (??) *Walks Nadia does her dog*.

Note: This grammar is so simple that it attaches *in the park* to *her dog* (i.e., Nadia owns a dog in the park and she walks it in some unspecified place), which is not how we would normally interpret this sentence in English. However, that doesn't affect the answer to this question.

Note: You do **not** need to consider sentences with modal verbs, such as *Should Nadia walk her dog in the park?*

Grammar:

S $\rightarrow$ NP VP
NP $\rightarrow$ N
NP $\rightarrow$ PN
NP $\rightarrow$ Det N
NP $\rightarrow$ Det Adj N
NP $\rightarrow$ NP PP
PP $\rightarrow$ P NP
VP $\rightarrow$ V
VP $\rightarrow$ V NP

Lexicon:

Det $\rightarrow$ *the* | *a* | *an* | *her* | *his*
Adj $\rightarrow$ *old* | *red* | *happy* | ...
N $\rightarrow$ *dog* | *old* | *park* | *statue* | *run* | ...
V $\rightarrow$ *saw* | *walk* | *ate* | *disdained* | ...
P $\rightarrow$ *in* | *to* | *on* | *under* | *with* | ...
PN $\rightarrow$ *Nadia* | *Ross* | ...

3. Playing with NLTK [6 marks]

Try out the interactive recursive-descent and shift-reduce parser demos in NLTK (`nltk.app.srparser()`, `nltk.app.rdparser()`), and answer the following questions.

a. With the shift-reduce parser and its default grammar, parse the sentence *the dog saw a man in the park with a statue* as many different ways as you can. Show the trees by printing out a screenshot of each one.

b. The default grammar with the shift-reduce parser includes the rule NP $\rightarrow$ NP PP, but that of the recursive-descent parser does not. Replace the NP rule of the latter with this rule:

   NP $\rightarrow$ NP PP | Det N

   and parse the default sentence. What happens? Now try:

   NP $\rightarrow$ Det N | NP PP

   Is this order change an adequate solution in general?

4. Bottom-up vs. top-down chart parsing [10 marks]

Consider the following sentence, given the grammar and lexicon below:

The best dogs like kids love fish.

Consider the agenda-based bottom-up and top-down chart parsing algorithms given in the lecture notes on chart parsing. If the agenda is treated as a stack (last-in-first-out) after initialization is complete, then what completed constituents would be built for this sentence by the bottom-up algorithm that would **not** be built by the top-down algorithm? Draw the tree structure of each such constituent, with words as the leaves of the trees.

Please arrange your trees in increasing order of size (number of nodes). This will make it easier for the grader to check that you have them all.

Hint: Be clear about what this question is asking for. You're being asked to parse the same sentence (by hand) by two different algorithms, and to observe that the bottom-up version builds some dead-end constituents that the top-down version is able to avoid.

Hint: This question is straightforward, but it takes some time to work through, so don't leave it until the last minute.

Grammar:

S $\rightarrow$ NP VP
NP $\rightarrow$ N
NP $\rightarrow$ Det N
NP $\rightarrow$ Det Adj N
NP $\rightarrow$ NP PP
PP $\rightarrow$ P NP
VP $\rightarrow$ V
VP $\rightarrow$ V NP

Lexicon (indexed by RHS):

the:	Det
best:	Adj, V
dogs:	N, V
like:	N, P, V
kids:	N, V
love:	N, V
fish:	N, V