

Assignment 2 – overview

CSC485/2501 – Fall 2019



UNIVERSITY OF
TORONTO

Most importantly

- This assignment is LONG
- Start working on question 3 and you'll see!

Question 1

- Getting your feet wet for Q3
- Update grammar to cover *imperatives* and *interrogatives*
- And nothing else!
- Regression testing
- Overspecification

S -> 'this' 'is' 'dumb'

Question 2

- For part a), count the number of *unique* parses you can make for the default sentence + grammar
 - Hint: more than one!
- For part b), edit the grammar (File > Edit Grammar), Autostep, and see what happens!

Question 3

*“Don’t be a tough guy[/gal].
Don’t be a fool!”*

- Donald President

Q3 – Getting Started

- Cannibalize rules from assignment sections 3.6.{1,2,3} and Chapter 12 of Jurafsky & Martin
 - Before asking on Blackboard, check J&M for an answer!
- Nonterminal to nonterminal rules (e.g. $S \rightarrow NP VP$) go in Grammar file
- Nonterminal to terminal rules (e.g. $V \rightarrow \text{'rewarded'}$) go in Lexicon file
- All parses will start with the S nonterminal
- You can use my scripts (see bulletin board) to run, check, and build ParseTrees, but *caveat emptor*
- Test-driven development

3.1 Very simple sentences

- Sentences with intransitive verbs in simple past tense
- Start with pronouns and verbs
 - she arrived
- Then get slightly more complex (determiners and nouns)
 - Nadia fondled the eggplant
- Then add modifiers (adverbs and adjectives)
 - Nadia left the handsome cat immediately
- Then prepositions
 - the cat with the long soft fur slowly ate

3.2 Auxiliaries

- Modal verbs + *be, have*
- Again, start with intransitive verbs
- Be careful of verb combinations and conjugations
 - *Is leaving, will leave, has left, have been leaving*
- Don't worry about auxiliary number agreement
 - The dogs has left

3.3 Subcategorization

- Verbs take different combinations of complements
- This depends on the verb itself
 - she arrived
 - Nadia fondled the eggplant
- This can be ungrammatical when left out
- You will use context-free rules, not features, to handle these differences
- There's nothing wrong with emulating them 😊

```
VP -> V  
VP -> V_NP NP  
V -> 'jumped'  
V_NP -> 'saw'
```

3.3 Subcategorization

- Sometimes one verb can take on different numbers of complements
 - Intransitive: Nadia ate
 - (Mono)transitive: Nadia ate the pie
- ... or different types of complements
 - NP: they told her a secret
 - Clause: they told her to go
- Such verbs should be listed multiple times in your lexicon

V → 'ate' | ...
V_NP → 'ate' | ...

Nouns and verbs

- Some words act as nouns and verbs
 - *Buffalo buffalo Buffalo buffalo buffalo buffalo Buffalo buffalo*
- List them multiple times in your lexicon with each role
- Maybe other words act in other grammatical roles

Testing

- Test on a number of your own sentences
 - Best if you do this while developing your grammar
 - The more the better!
- Try generating random sentences from your rules
- Describe and give examples of:
 - Overgeneration – parsed when it shouldn't have
 - Undergeneration – didn't parse when it should have
- I will test on a bunch of private sentences
 - Similar to those given in the assignment (no new grammatical constructs, but likely different verb conjugations)
- **If your grammar can't be parsed by NLTK on CDF, you will receive a zero!!!!!!**