Syracuse University

IST-664 NLP Homework3
Writing a CFG (Context-Free Grammar)

ThulasiRam RuppaKrishnan
IST 664
Professor Michael Larche

Contents

Introduction

For this homework assignment, develop a CFG for a small subset of English. There will be just one part, writing grammar rules, with no option for programming.

The grammar rules will be written in a file that can be read by the NLTK. The first part of the grammar file contains the structure rules for the phrases and the second part contains the lexical rules that assign non-terminal symbols for the words in the vocabulary of this limited language. These non-terminal symbols can be thought of as POS tags, but are not required to be.

There will be a set of sentences to try to parse, given in a file called sentences.txt. There will be a second file of sentences, challenge_sentences.txt, from which we need to choose two more sentences to parse.

About the data and Programme

Download the CFGdevelopment.zip file and extract the directory that has the following structure:

CFGdevelopment/
- camelot_grammar.cfg
- challenge_sentences.txt
- sentences.txt
- development.py

The file development.py has a small bit of python that reads sentences from the file sentences.txt, reads the grammar from camelot_grammar.cfg, produces a recursive descent parser and parses the sentences, if possible. When you choose two challenge sentences, you can copy them to the end of the file sentences.txt.

You will add grammar rules to the file camelot_grammar.cfg to parse the sentences. The NLTK program that loads grammars from .cfg files will ignore blank lines and comment lines starting with #.

Writing Context Free Grammers

New rules and Lexical definitions

There are numerous new rules and lexical definitions were added in the camelot_grammar.cfg to parse all the original and two challenge sentences which are available in the following attached file

Part 1: original sentences in sentences.txt and two challenge sentences.

Original sentences

1. Arthur is the king .

   ['Arthur', 'is', 'the', 'king', '.']
   (START
     (S1
       (NP (Proper Arthur))
       (VP (VerbT is) (NP (Det the) (NP (Noun king))))
       (Eos .)))
   (START
     (S1
       (NP (Proper Arthur))
       (VP (VerbT is) (NP (Det the) (Noun king)))
       (Eos .)))

   This sentence comprises of a Noun Phrase and a Verb Phrase where the verb phrase has “is” verb and a noun phrase. No new rules were needed in the grammar file to parse this sentence

2. Arthur rides the horse near the castle .

   ['Arthur', 'rides', 'the', 'horse', 'near', 'the', 'castle', '.']
   (START
     (S1
       (NP (Proper Arthur))
       (VP
         (VerbT rides)
         (NP

(Det the)
(NP
(Noun horse)
(PP (Prep near) (NP (Det the) (NP (Noun castle)))))))
(Eos .)))
(START
(S1
(NP (Proper Arthur))
(VP
(VerbT rides)
(NP (Det the) (NP (Noun horse)))
(PP (Prep near) (NP (Det the) (NP (Noun castle)))))
(Eos .)))

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of a Noun Phrase and a Prepositional Phrase. No new rules were added to the grammar file in this case

3. Arthur rides the plodding horse near the castle .

['Arthur', 'rides', 'the', 'plodding', 'horse', 'near', 'the', 'castle', '.']
(START
(S1
(NP (Proper Arthur))
(VP
(VerbT rides)
(NP (Det the) (JJ plodding) (Noun horse))
(PP (Prep near) (NP (Det the) (NP (Noun castle)))))
(Eos .)))
(START
(S1
(NP (Proper Arthur))
(VP
(VerbT rides)
(NP (Det the) (JJ plodding) (Noun horse))
(PP (Prep near) (NP (Det the) (Noun castle))))
(Eos .)))

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of a Noun Phrase and a Prepositional Phrase. New rules were added to the grammar file to parse the token “plodding” as adjective and included as part of the Noun Phrase definition

4. the Holy_Grail is a chalice .

['the', 'Holy_Grail', 'is', 'a', 'chalice', '.']
(START

```
  (S1
    (NP (Det the) (NP (NNP Holy_Grail)))
    (VP (VerbT is) (NP (Det a) (NP (Noun chalice))))
    (Eos .)))
(START
  (S1
    (NP (Det the) (NP (NNP Holy_Grail)))
    (VP (VerbT is) (NP (Det a) (Noun chalice)))
    (Eos .)))
(START
  (S1
    (NP (Det the) (NNP Holy_Grail))
    (VP (VerbT is) (NP (Det a) (NP (Noun chalice))))
    (Eos .)))
(START
  (S1
    (NP (Det the) (NNP Holy_Grail))
    (VP (VerbT is) (NP (Det a) (Noun chalice)))
    (Eos .)))
```

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of “is” verb and a Noun Phrase. New rules were added to the grammar file to parse the token “Holy_Grail” as noun (NNP) which is not defined in proper noun lexical rule. This is also now, included as part of the Noun Phrase definition

5. the sensational Holy_Grail is a sacred chalice .

```
['the', 'sensational', 'Holy_Grail', 'is', 'a', 'sacred', 'chalice', '.']
(START
  (S1
    (NP (Det the) (JJ sensational) (NNP Holy_Grail))
    (VP (VerbT is) (NP (Det a) (JJ sacred) (Noun chalice)))
    (Eos .)))
```

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of “is” verb and a Noun Phrase. New rules were added to the grammar file to include adjectives (JJ) as part of the Noun Phrase definition with NNP and Noun lexical.

6. every coconut was carried to the hottest mountains .

```
['every', 'coconut', 'was', 'carried', 'to', 'the', 'hottest', 'mountains', '.']
(START
  (S1
    (NP (Det every) (NP (Noun coconut)))
    (VP
```

```
    (Vbar was (VBpast carried))
    (TO to)
    (NP (Det the) (JJS hottest) (NNS mountains)))
   (Eos .)))
(START
 (S1
  (NP (Det every) (Noun coconut))
  (VP
   (Vbar was (VBpast carried))
   (TO to)
   (NP (Det the) (JJS hottest) (NNS mountains)))
  (Eos .)))
```

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of an Auxiliary verb “was” along with the main verb “carried” then with “TO” and a Noun Phrase. New lexical definition is added to define superlative adjectives as JJS and plural nouns as NNS. New rules were added to the grammar file to include adjectives (JJS) as part of the Noun Phrase definition with NNS.

7. sixty strangers are at the Round_Table .

['sixty', 'strangers', 'are', 'at', 'the', 'Round_Table', '.']

```
(START
 (S1
  (NP (CD sixty) (NNS strangers))
  (VP
   (VBP are)
   (PP (Prep at) (NP (Det the) (NP (NNP Round_Table)))))
  (Eos .)))
(START
 (S1
  (NP (CD sixty) (NNS strangers))
  (VP (VBP are) (PP (Prep at) (NP (Det the) (NNP Round_Table))))
  (Eos .)))
```

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of a plural verb “are and a Prepositional Phrase. New lexical definition is added to define “Round_Table” as NNP and “sixty” as CD. New rules were added to the grammar file to include count determiners (CD) as part of the Noun Phrase definition with NNS.

8. Sir_Lancelot might have spoken .

```
['Sir_Lancelot', 'might', 'have', 'spoken', '.']
(START
```

```
(S1
  (NP (Proper Sir_Lancelot))
  (VP (MD might) (VB have) (VBN spoken))
  (Eos .)))
```

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of a modal “might”, a base form verb like “have” and a past participle verb “spoken”. New lexical definitions are added to define “might” as MD and past participle verbs as VBN. New rules were added to the grammar file to include the combination of MD VB and VBN as a verb Phrase.

9. Guinevere had been riding with Patsy for five weary nights .

```
['Guinevere', 'had', 'been', 'riding', 'with', 'Patsy', 'for', 'five', 'weary', 'nights', '.']
(START
  (S1
    (NP (Proper Guinevere))
    (VP
      (Vbar (VBpast had) (VBN been) (VBG riding))
      (PP (Prep with) (NP (Proper Patsy)))
      (PP (Prep for) (NP (CD five) (JJ weary) (NNS nights))))
    (Eos .)))
```

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of a past form verb “had”, “been” verb and a continuous verb “riding” along with two Preposition Phrase. New lexical definitions are added to define “had” as VBpast and continuous verb form as VBG. New rules were added to the grammar file to include the combination of VBpast VBN and VBG as Vbar and combination Vbar with PP as a verb Phrase.

10. Sir_Bedevere might have been suggesting this quest .

```
['Sir_Bedevere', 'might', 'have', 'been', 'suggesting', 'this', 'quest', '.']
(START
  (S1
    (NP (Proper Sir_Bedevere))
    (VP
      (Vbar (MD might) (VB have) (VBN been) (VBG suggesting))
      (NP (Det this) (NP (Noun quest))))
    (Eos .)))
(START
  (S1
    (NP (Proper Sir_Bedevere))
    (VP
      (Vbar (MD might) (VB have) (VBN been) (VBG suggesting))
```

```
    (NP (Det this) (Noun quest)))
  (Eos .)))
```

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of a modal “might”, verb “have”, “been” verb and a continuous verb “suggesting” along with a Noun Phrase. New rules were added to the grammar file to include the combination of MD, VB, VBN and VBG as Vbar and combination Vbar with NP as a verb Phrase.

11. the Britons migrate south frequently .

```
['the', 'Britons', 'migrate', 'south', 'frequently', '.']
(START
  (S1
    (NP (Det the) (NNPS Britons))
    (VP (VB migrate) (RB south) (RB frequently))
    (Eos .)))
```

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of a verb and an adverb. New Lexical definitions were added to treat “south” and “frequently” as adverbs (RB). New rules were added to the grammar file to include the combination of VB, RB, RB as a verb Phrase.

12. Arthur and Guinevere ride frequently near the castle .

```
['Arthur', 'and', 'Guinevere', 'ride', 'frequently', 'near', 'the', 'castle', '.']
(START
  (S1
    (NP (Proper Arthur) (CC and) (Proper Guinevere))
    (VP
      (VB ride)
      (RB frequently)
      (PP (Prep near) (NP (Det the) (NP (Noun castle)))))
    (Eos .)))
(START
  (S1
    (NP (Proper Arthur) (CC and) (Proper Guinevere))
    (VP
      (VB ride)
      (RB frequently)
      (PP (Prep near) (NP (Det the) (Noun castle))))
    (Eos .)))
```

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of a verb , an adverb and a preposition phrase .Also, Noun phrase consists of two proper nouns and a conjunction (CC). New Lexical definitions were added to treat “and” as a

conjunction (CC). New rules were added to the grammar file to include the combination of Proper, CC and Proper as a Noun phrase and VB RB PP as a Verb Phrase.

13. he suggests to grow fruit at home .

```
['he', 'suggests', 'to', 'grow', 'fruit', 'at', 'home', '.']
(START
  (S1
    (NP (PRP he))
    (VP
      (Vbar (VBZ suggests) (TO to) (VB grow))
      (NP (Noun fruit) (PP (Prep at) (NP (Noun home)))))
    (Eos .)))
(START
  (S1
    (NP (PRP he))
    (VP
      (Vbar (VBZ suggests) (TO to) (VB grow))
      (NP (Noun fruit))
      (PP (Prep at) (NP (Noun home))))
    (Eos .)))
(START
  (S1
    (NP (PRP he))
    (VP
      (Vbar (VBZ suggests) (TO to) (VB grow))
      (NP (Noun fruit) (PP (Prep at) (NP (Noun home)))))
    (Eos .)))
(START
  (S1
    (NP (PRP he))
    (VP
      (Vbar (VBZ suggests) (TO to) (VB grow))
      (NP (Noun fruit))
      (PP (Prep at) (NP (Noun home))))
    (Eos .)))
```

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of a third person singular verb, “TO” and a verb “VB”. Also, Noun phrase consists of personal pronouns (PRP). New Lexical definitions were added PRP and third person singular verb VBZ. New rules were added to the grammar file to include the combination of VBZ, TO and VB as Vbar Also, the combination of Vbar NP and PP as Verb Phrase.

14. riding to Camelot is not hard .

['riding', 'to', 'Camelot', 'is', 'not', 'hard', '.']
(START
  (S1
    (NP (VBG riding))
    (VP (Vbar (TO to) (NNP Camelot) (VerbT is) (NOT not) (JJ hard)))
    (Eos .)))

This sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of “TO”, NNP, “is” verb, “NOT” and an adjective “JJ”. Also, Noun phrase consists of personal pronouns (VBG). New rules were added to the grammar file to include the combination of TO, NNP, VerbT, NOT and JJ as Vbar which is a verb phrase and just VBG as a noun phrase

15. do coconuts speak ?

['do', 'coconuts', 'speak', '?']
(START (S1 (DO do) (NP (NNS coconuts)) (VP (VB speak)) (Eos ?)))

This sentence comprises of “DO”, Noun Phrase and a Verb Phrase where the Verb Phrase consists of a verb VB. Also, Noun phrase consists of NNS. New rules were added to the grammar file to include the combination of DO, NP , VP and Eos as a sentence structure.

16. why does England have a king ?

['why', 'does', 'England', 'have', 'a', 'king', '?']
(START
  (S1
    (WRB why)
    (DO does)
    (NP (NNP England))
    (VP (VB have) (NP (Det a) (NP (Noun king))))
    (Eos ?)))
(START
  (S1
    (WRB why)
    (DO does)
    (NP (NNP England))
    (VP (VB have) (NP (Det a) (Noun king)))
    (Eos ?)))

This sentence comprises of wh adverb “WRB”, “DO”, Noun Phrase and a Verb Phrase where the Verb Phrase consists of a verb VB and a NP. Also, Noun phrase consists of NNP. New rules were added to the grammar file to include the combination of WRB, DO, NP, VP and Eos as a sentence structure.

Challenge sentences

1. it is Sir_Lancelot who knows Zoot !

   ['it', 'is', 'Sir_Lancelot', 'who', 'knows', 'Zoot', '!']
   (START
     (S1
       (NP (PRP it))
       (VP (VerbT is) (NP (Proper Sir_Lancelot)))
       (WP who)
       (VP (VBZ knows) (NP (Proper Zoot)))
       (Eos !)))
   (START
     (S1
       (NP (PRP it))
       (VP (VerbT is) (NP (Proper Sir_Lancelot)))
       (WP who)
       (VP (VBZ knows) (NP (Proper Zoot)))
       (Eos !)))

   This challenge sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of two verb phrase and wh pronouns WP. Also, Noun phrase consists of PRP. New rules were added to the grammar file to include the combination of NP, VP, WP, VP and Eos as a sentence structure.

2. either Arthur knows or Patsy does .

   ['either', 'Arthur', 'knows', 'or', 'Patsy', 'does', '.']
   (START
     (S1
       (NP (CC either) (Proper Arthur))
       (VP (VBZ knows) (CC or) (NP (Proper Patsy)) (DO does))
       (Eos .)))

   This challenge sentence comprises of Noun Phrase and a Verb Phrase where the Verb Phrase consists of VBZ, CC, NP and DO. Also, Noun phrase consists of CC and Proper noun. New rules were added to the grammar file to include the combination of CC with a proper noun as noun phrase also VBZ, CC, NP and DO as a verb phrase.

Part 2: Exemplar sentences for no parse and overgeneralization sentences with parse.

Exemplar sentences with no parse

1. suggesting horse ride near the hottest mountains for strangers is sensational .

   This sentence cannot be parsed with the existing grammar rules and lexical definitions because it requires new rule for parsing two nouns NN NN with VBG as noun phrase and combination PP PP and a verb phrase (VP) as VP even though the lexical has already occurred in the sentence which was parsed successful.

   suggesting_VBG horse_NN ride_NN near_Prep the_DT hottest_JJS mountains_NNS for_Prep strangers_NNS is_VBZ sensational_JJ

2. he knows when to migrate and grow fruit near south .

   This sentence cannot be parsed with the existing grammar rules and lexical definitions because it requires new rule for parsing VBZ , WRB, TO, VB, CC , VB, NN, Prep and RB as a verb Phrase and just PRP as a noun phrase even though the lexical has already occurred in the sentence which was parsed successful.

   he_PRP knows_VBZ when_WRB to_TO migrate_VB and_CC grow_VB fruit_NN near_Prep south_RB .

3. why does the Camelot horse ride for five weary nights ?

   This sentence cannot be parsed with the existing grammar rules and lexical definitions because it requires new rule for parsing WRB and verb phrase consists of VBZ DT NNP NN NN Prep CD JJ and NNS even though the lexical has already occurred in the sentence which was parsed successful.

   why_WRB does_VBZ the_DT Camelot_NNP horse_NN ride_NN for_Prep five_CD weary_JJ nights_NNS ?_.

Overgeneralization sentences with parse

1. Holy_Grail grow at Arthur .

```
['Holy_Grail', 'grow', 'at', 'Arthur', '.']
(START
  (S1
    (NP (NNP Holy_Grail))
    (VP (VBP grow) (PP (Prep at) (NP (Proper Arthur))))
    (Eos .)))
```

This is not an English sentence where VBP “grow” with preposition “at ” cannot occur before a proper noun but still our rules are very generalized to parse this successfully.

2. Guinevere were been drinking across Dingo of sixty further ants .

```
['Guinevere', 'were', 'been', 'drinking', 'across', 'Dingo', 'of', 'sixty', 'further', 'ants', '.']
(START
  (S1
    (NP (Proper Guinevere))
    (VP
      (Vbar (VBpast were) (VBN been) (VBG drinking))
      (PP (Prep across) (NP (Proper Dingo)))
      (PP (Prep of) (NP (CD sixty) (JJ further) (NNS ants))))
    (Eos .)))
```

This is not an English sentence where “were” “been” “drinking” cannot occur in the sequence but still our rules are very generalized to parse this successfully.

3. this Holy_Grail covers this trusty defeater .

```
['this', 'Holy_Grail', 'covers', 'this', 'trusty', 'defeater', '.']
(START
  (S1
    (NP (Det this) (NP (NNP Holy_Grail)))
    (VP (VerbT covers) (NP (Det this) (JJ trusty) (Noun defeater)))
    (Eos .)))
(START
  (S1
    (NP (Det this) (NNP Holy_Grail))
    (VP (VerbT covers) (NP (Det this) (JJ trusty) (Noun defeater)))
    (Eos .)))
```

This is not an English sentence where NNP “Holy_Grail” cannot come before covers which cannot occur before “this trusty defeater” in the sequence but still our rules are over generalized to parse this successfully.