Homework 3 – Context Free Grammar

Joyce LJ Woznica

Table of Contents

[Introduction 1](#_Toc49003782)

[Description 1](#_Toc49003783)

[Part 1 – Writing Grammar Rules 2](#_Toc49003784)

[Starting Grammar 2](#_Toc49003785)

[Standardizing the Part of Speech Tags 3](#_Toc49003786)

[Changes to Starting Sentences 4](#_Toc49003787)

[Parsing of Sentences 5](#_Toc49003788)

[Sentence 3 5](#_Toc49003789)

[Sentence 4 5](#_Toc49003790)

[Sentence 5 6](#_Toc49003791)

[Sentence 6 6](#_Toc49003792)

[Sentence 7 7](#_Toc49003793)

[Sentence 8 8](#_Toc49003794)

[Sentence 9 8](#_Toc49003795)

[Sentence 10 9](#_Toc49003796)

[Sentence 11 9](#_Toc49003797)

[Sentence 12 10](#_Toc49003798)

[Sentence 13 10](#_Toc49003799)

[Sentence 14 11](#_Toc49003800)

[Sentence 15 12](#_Toc49003801)

[Sentence 16 12](#_Toc49003802)

[Challenge Sentence 1 13](#_Toc49003803)

[Challenge Sentence 2 13](#_Toc49003804)

[Part 2 – Exemplar Sentences 14](#_Toc49003805)

[Sentence That Cannot be Parsed (Overfitting) 14](#_Toc49003806)

[Gibbersh Sentence that Parses (Overgeneralization) 14](#_Toc49003807)

[Conclusion 15](#_Toc49003808)

# Introduction

The objective of this homework is develop a context free grammar (CFG) for a small subset of English particularly – those found in the movie *Monty Python and the Holy Grail*. All vocabulary to be used for this assignment is already provided in the configuration file.

# Description

The grammar exists in a configuration file that is read by a provided python code file. There are sentences in a separate file. The goal is to properly parse all the sentences in this file and add at least two sentences from an additional file and parse them as well.

Finally, we will look at overgeneralization and over fitting and discuss the reasons why this happens with the CFG provided. The first two sentences are parsed for us so that we can use this information as an example.

# Part 1 – Writing Grammar Rules

In this section, additional rules so that all sentences can be parsed. Each sentence and the added changes are noted as well as the result and explanation.

## Starting Grammar

Initial context free grammar was provided in the figure below:

A screenshot of a social media post

Description automatically generated

Figure Initial Grammar

This initial grammar parsed the sentences below as follows:

Arthur is the King .

Parsed as follows:

(START

(S1

(NP (Proper Arthur))

(VP (VerbT is) (NP (Det the) (NP (Noun king))))

(Eos .)))

The second sentence was:

Arthur rides the horse near the castle .

Parsed as follows:

['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)))))

### Standardizing the Part of Speech Tags

Modifications of the provided vocabulary tags were done using the Penn Treebank standard. The following changes were made to the provided vocabulary:

* *Noun* was changed to ***NN*** (noun, singular or mass)
* *Det* was changed to ***DT*** (determiner)
* *Prep* was changed to ***IN*** (preposition)
* *Proper* was changed to ***NNP*** (proper noun, singular)
* *VerbT* was changed the ***VBZ*** (verb, 3rd person singular, present)
* *Misc1* was changed to ***PUNC*** (my tag for punctuation)
* *Misc2* was changed to ***CC*** (coordinating conjunction)
* *Misc3* was changed to ***CD*** (cardinal number)
* *Misc4* was changed to ***EX*** (existential there)
* *Misc5* was changed to ***CCS*** (my code for subordinating conjunctions)
* *Misc6* was changed to ***MD*** (modals)
* *Misc7* was changed to ***JJ*** (adjectives)
* *Misc8* was changed to ***JJR*** (comparative adjectives)
* *Misc9* was changed to ***JJS*** (superlative adjectives)
* *Misc10* was changed to ***NNS*** (noun, plural)
* *Misc11* was changed to ***NNO*** (my code for a noun object or place)
* *Misc12* was changed to ***NNPS*** (proper noun, plural)
* *Misc13* was changed to ***PRP*** (personal pronoun)
* *Misc14* was changed to ***PRP$*** (possessive pronoun)  
  Note – this errors in the python code, so for this exercise ***PRPS*** was used due to errors in the code when a “$” is introduced.
* *Misc15* was changed to ***RB*** (adverb)
* *Misc16* was changed to ***VB*** (verb, base form)
* *Misc17* was changed to ***VBD*** (verb, past tense)
* *Misc18* was changed ***VBG*** (verb, gerund/present participle)
* *Misc19* was changed to ***VBN*** (verb, past participle)
* *Misc20*  was changed to ***VBZ2***(my code for additional 3rd person singular verbs, present)
* *Misc21* was changed to ***VBP*** (3rd person plural verbs, present)
* *Misc22* was changed to ***WDT*** (wh-determiner)
* *Misc23* was changed to ***WP*** (wh-pronoun)
* *Misc24* was changed to ***WPS*** (possessive wh-pronoun)
* *Misc25* was changed to ***WRB*** (wh-adverb)
* *POSS* was changed to ***POS*** (possessive ending)

The following tags were not modified:

* *Eos*
* *DO*
* *NOT*
* *TO*

For these modifications, I reviewed the following for the Penn Treebank Part of Speech Tags found in the following websites:

* <http://www.surdeanu.info/mihai/teaching/ista555-all13/readings/PennTreebankConstituents.html>. This list was selected because it was the link referenced in our course materials.
* <https://www.cs.upc.edu/~nlp/SVMTool/PennTreebank.html>. I selected this to assist because it provided examples of each tag which I found very helpful.

As shown, I created a couple tags of my own where I could not find an existing tag.

As recommended in class, the following lines were modified as well in the provided grammer:

# Noun phrases

NP -> DT NP | NNP | NN PP | NN

This initial grammar was changed to replace the determiner followed by a noun phrase (DT NP) to simply a determiner followed by a noun (DT NN):

# maybe change to

NP -> DT NN | NNP | NN PP | NN

### Changes to Starting Sentences

Making these modifications, changed the initial tagging of the first two sentences as follows:

#### Sentence 1

Arthur is the King .

Parsed as follows:

['Arthur', 'is', 'the', 'king', '.']

(START

(S1

(NP (NNP Arthur))

(VP (VBZ is) (NP (DT the) (NN king)))

(Eos .)))

#### Description

This sentence is made up of an NP (noun phrase) “Arthur” followed by a VP (verb phrase) “is the king” and then an Eos (End of Sentence) “.”. Each of these phrases is further parsed with the noun phrase then tagging “Arthur” as a NNP (proper noun) and the verb phrase parsing into VBZ (3rd person singular verb) “is” followed by another NP (noun phrase) “the king”. The final noun phrase is parsed to a DT (determiner) “the” followed by a NN (noun) “king” and finally the Eos “.” for the end of the sentence.

#### Sentence 2

The second sentence was:

Arthur rides the horse near the castle .

Parsed as follows:

['Arthur', 'rides', 'the', 'horse', 'near', 'the', 'castle', '.']

(START

(S1

(NP (NNP Arthur))

(VP

(VBZ rides)

(NP (DT the) (NN horse))

(PP (IN near) (NP (DT the) (NN castle))))

(Eos .)))

#### Description

This sentence was parsed as a noun phrase (NP) which is an NNP (proper noun) “Arthur” followed by a verb phrase (VP) which is then parsed as a VBZ (a present tense 3rd person verb) “rides” followed by another noun phrase (NP) . This next NP is parsed as a DT (determiner) “the” and a NN (noun) “horse” then a preposition phrase (PP) made up the IN (preposition) “near” and of a NP (noun phrase) which is a DT (determiner) “the” and a noun phrase of a single NN (noun) “castle” and then an Eos (end of sentence) “.”.

## Parsing of Sentences

Now that all the tagging is standard, the exercise of adding to the grammar begins. Original grammar in the *camelot\_grammar.cfg* file is noted in blue font. Changes made to handle parsing of the new sentences are noted in **green bold font**.

### Sentence 3

Arthur rides the plodding horse near the castle .

In reviewing this sentence, we notice that there is a noun phrase “Arthur” followed by a verb phrase “rides the plodding horse near the castle”. This verb phrase includes the verb “rides” then a determiner “the” followed by an adjective “plodding” and then another determiner “the” followed by the preposition “near”, so the following was changed to properly handle this sentence by adding a noun phrase which consists of a DT (determiner) followed by an JJ (adjective) and a NN (noun).

# Noun phrases

NP -> DT NN | NNP | NN PP | NN **| DT JJ NN**

This change parsed the sentence as follows:

['Arthur', 'rides', 'the', 'plodding', 'horse', 'near', 'the', 'castle', '.']

(START

(S1

(NP (NNP Arthur))

(VP

(VBZ rides)

(NP (DT the) (JJ plodding) (NN horse))

(PP (IN near) (NP (DT the) (NN castle))))

(Eos .)))

#### Description

This sentence consists of a NP (noun phrase) “Arthur” followed by a VP (verb phrase) “rides the plodding horse near the castle”. The first NP is just a proper noun NNP “Arthur” and the verb phrase is a VBZ (3rd person present tense verb) “rides” followed by another NP (noun phrase). The latter noun phrase is then a DT (determiner) “the” followed by a JJ (adjective) “plodding” followed by an NN (noun) “horse” then followed by a PP (prepositional phrase) which is a IN (preposition) “near” followed by yet another NP (noun phrase) made up of a DT (determiner) “the” followed by a NN (noun) “castle” and an Eos (end of sentence) “.”.

### Sentence 4

the Holy\_Grail is a chalice .

In reviewing this sentence, we notice that there is a noun phrase “the Holy\_Grail” followed by a verb phrase “is a chalice”. This verb phrase includes a verb “is” then a determiner “a” followed by noun “chalice”, so the following was changed to properly handle this sentence by adding a noun phrase which is a DT (determiner) followed by an NNO (my tag for a noun or a proper object or place):

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN **| DT NNO**

This change parsed the sentence as follows:

['the', 'Holy\_Grail', 'is', 'a', 'chalice', '.']

(START

(S1

(NP (DT the) (NNO Holy\_Grail))

(VP (VBZ is) (NP (DT a) (NN chalice)))

(Eos .)))

#### Description

This sentence consists of an NP (noun phrase) “the Holy\_Grail” followed by a VP (verb phrase) “is a chalice”. The first noun phrase is a DT (determiner) “the” followed by an NNO (noun, object or place) “Holy\_Grail”. The VP (verb phrase) includes a VBZ (3rd person present tense verb) “is” then a DT (determiner) “a” followed by NN (noun) “chalice” finished with an Eos (end of sentence) ”.”.

### Sentence 5

the sensational Holy\_Grail is a sacred chalice .

In reviewing this sentence, we notice that there is a noun phrase “the sensational Holy\_Grail” followed by a verb phrase “is a sacred chalice”. The noun phrase is not covered by our existing grammar and we need to add a DT (determiner) followed by an JJ (adjective) and then an NNO (noun for proper object or place). This verb phrase includes a verb “is” then a determiner “a” followed by an adjective “sacred” then the noun “chalice”, so the following was changed to properly handle this sentence:

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO **| DT JJ NNO**

This change parsed the sentence as follows:

['the', 'sensational', 'Holy\_Grail', 'is', 'a', 'sacred', 'chalice', '.']

(START

(S1

(NP (DT the) (JJ sensational) (NNO Holy\_Grail))

(VP (VBZ is) (NP (DT a) (JJ sacred) (NN chalice)))

(Eos .)))

#### Description

This sentence consists of a NP (noun phrase) “the sensational Holy\_Grail” which is a DT (determiner) “the” followed by a JJ (adjective) and then a NNO (proper object or place noun). This noun phrase is followed by a VP (verb phrase) which consists of a VBZ (3rd person singular verb) “is” and another NP (noun phrase) which is a DT (determiner) “a” followed by a JJ (adjective) “sacred” and a NN (noun) “chalice” with an Eos (end of sentence) “.”.

### Sentence 6

every coconut was carried to the hottest mountains .

In reviewing this sentence, we notice that there is a noun phrase “every coconut” followed by a verb phrase “was carried to the hottest mountains”. The noun phrase is a determiner followed by a noun. This verb phrase includes a past tense verb “was” then a past participle verb “carried” followed by a noun phrase. The noun phrase is the preposition “to”, which makes it a preposition phrase followed by a noun phrase, a determiner “the”, an adjective (superlative) “hottest” and a plural noun “mountains”.

The following was changed to properly handle this sentence:

# Verb phrases

VP -> VBZ NP | VBZ NP PP **| VBD VBN NP**

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO **| TO NP | DT JJS NNS**

This change parsed the sentence as follows:

['every', 'coconut', 'was', 'carried', 'to', 'the', 'hottest', 'mountains', '.']

(START

(S1

(NP (DT every) (NN coconut))

(VP

(VBD was)

(VBN carried)

(NP (TO to) (NP (DT the) (JJS hottest) (NNS mountains))))

(Eos .)))

#### Description

The sentence is parsed as an NP (noun phrase) followed by a VP (verb phrase). The NP is a DT (determiner) “every” followed by a single NN (noun) “coconut”. The VP (verb phrase) is a VBD (past tense verb) “was” followed by a VBN (past participle verb) “carried” followed by another NP which consists of TO (the word “to”) and a NP of a DT (determiner) “the” followed by a JJS (superlative adjective) “hottest” and then a NNS (a plural noun) “mountains” and an Eos (end of sentence) “.”.

### Sentence 7

sixty strangers are at the Round\_Table .

In reviewing this sentence, we have a noun phrase of a cardinal number “sixty” and a plural noun “strangers” followed by a verb phrase with a plural verb “are”, a preposition “at”, a determiner “the” and a proper noun location or place “Round\_Table”. Adding a new verb phrase of a VBP (plural verb) followed by a PP (preposition phrase) and then a new noun phrase of a CD (cardinal number) followed by a NNS (plural noun ) had to be added to properly handle this sentence:

# Verb phrases

VP -> VBZ NP | VBZ NP PP | VBD VBN NP | **VBP PP**

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO | TO NP | DT JJS NNS **| CD NNS**

This change parsed the sentence as follows:

['sixty', 'strangers', 'are', 'at', 'the', 'Round\_Table', '.']

(START

(S1

(NP (CD sixty) (NNS strangers))

(VP (VBP are) (PP (IN at) (NP (DT the) (NNO Round\_Table))))

(Eos .)))

#### Description

This sentence is parse as a NP (noun phrase) which is a CD (cardinal number) “sixty” followed by an NNS (plural noun) “strangers” followed by a VP (verb phrase) of the VBP (plural 3rd person verb) and a PP (preposition phrase) of a IN (preposition) “at” followed by an NP (noun phrase) of a DT (determiner) “the” followed by an NNO (proper noun object) followed by an Eos (end of sentence) “.”.

### Sentence 8

Sir\_Lancelot might have spoken .

In reviewing this sentence, we have a noun phrase of a proper noun “Sir\_Lancelot” followed by a verb phrase that begins with a modal verb “might” and then a base verb “have” and then a past tense verb “spoken”.

The following was changed to properly handle this sentence:

# Verb phrases

VP -> VBZ NP | VBZ NP PP | VBD VBN NP | VBP PP **| MD VB VBN**

This change parsed the sentence as follows:

['Sir\_Lancelot', 'might', 'have', 'spoken', '.']

(START

(S1

(NP (NNP Sir\_Lancelot))

(VP (MD might) (VB have) (VBN spoken))

(Eos .)))

#### Description

This sentence was parsed as a NP (noun phrase) which consists of a single NNP (proper noun) “Sir\_Lancelot” which is followed by a VP (verb phrase) consisting of a MD (modal) “might” then a VB (base verb) and a VBN (past participle verb) “spoken” then an Eos (end of sentence) “.”.

### Sentence 9

Guinevere had been riding with Patsy for five weary nights .

In reviewing this sentence, we have a noun phrase of a proper noun “Guinevere” followed by a verb phrase that begins with past tense verb “had” and then a past participle verb ”been” then a present participle verb “riding” and a preposition phrase “with Patsy for five weary nights”. The preposition phase is a preposition “with” followed by a noun phrase which is a proper noun “Patsy” and another preposition phrase of “for five weary nights.” The final preposition phrase is a preposition “for” and a noun phrase of a cardinal number “five” an adjective “weary” and a plural noun “nights”.

The following was changed to properly handle this sentence:

# Verb phrases

VP -> VBZ NP | VBZ NP PP | VBD VBN NP | VBP PP | MD VB VBN **| VBD VBN VBG PP**

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO | TO NP | DT JJS NNS | CD NNS **| NNP PP | CD JJ NNS**

This change parsed the sentence as follows:

['Guinevere', 'had', 'been', 'riding', 'with', 'Patsy', 'for', 'five', 'weary', 'nights', '.']

(START

(S1

(NP (NNP Guinevere))

(VP

(VBD had)

(VBN been)

(VBG riding)

(PP

(IN with)

(NP

(NNP Patsy)

(PP (IN for) (NP (CD five) (JJ weary) (NNS nights))))))

(Eos .)))

#### Description

This sentence was parsed into a NP (noun phrase) which consists of just a NNP (proper noun) “Guinevere”. This is then followed by a VP (verb phrase) which consists of a VBD (past tense verb) “had” followed by a VBN (past participle verb) ”been” then a VBG (present participle verb) “riding”. The next part of the verb phrase is a PP (preposition phrase) “with Patsy for five weary nights”. This parses to an IN (preposition) “with” followed by another NP (noun phrase) which consists of a NNP (a proper noun) “Patsy” followed by another PP ( preposition phrase) “for five weary nights” which is an IN (preposition) “for” followed by a CD (cardinal number) “five”, a JJ (adjective) “weary” and then a NNS (plural noun) “nights” ending with an Eos (end of sentence) “.”.

### Sentence 10

Sir\_Bedevere might have been suggesting this quest .

In reviewing this sentence, we see that there is longer verb phrase following the noun phrase “Sir\_Bedevere” which consists of a modal “might” then a base verb “have”, a past participle “been” and then a present participle verb “suggesting” followed by a noun phrase “this quest”.

The following was changed to properly handle this sentence:

# Verb phrases

VP -> VBZ NP | VBZ NP PP | VBD VBN NP | VBP PP | MD VB VBN | VBD VBN VBG PP **| MD VB VBN VBG NP**

This change parsed the sentence as follows:

['Sir\_Bedevere', 'might', 'have', 'been', 'suggesting', 'this', 'quest', '.']

(START

(S1

(NP (NNP Sir\_Bedevere))

(VP

(MD might)

(VB have)

(VBN been)

(VBG suggesting)

(NP (DT this) (NN quest)))

(Eos .)))

#### Description

This sentence is parsed as a NP (noun phrase) which is a single NNP (proper noun) “Sir\_Bedevere” followed by a VP (verb phrase). The VP is a MD (model) “might” followed by a VB (base verb) “have” then a VBN (past participle verb) “been” and a VBG (present participle verb) “suggesting” followed by another NP (noun phrase). The final NP is a DT (determiner) “this” followed by a single NN (noun) “quest” and then an Eos (End of sentence) “.”.

### Sentence 11

the Britons migrate south frequently .

In reviewing this sentence, we see that we need to have a new noun phrase to cover a determiner followed by a plural proper noun for “the Britons” and that there is new verb phrase grammar required to cover a base verb followed by not one, but two adverbs together ”migrate south frequently”, so the following was changed to properly handle this sentence:

# Verb phrases

VP -> VBZ NP | VBZ NP PP | VBD VBN NP | VBP PP | MD VB VBN | VBD VBN VBG PP | MD VB VBN VBG NP **| VB RB RB**

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO | TO NP | DT JJS NNS | CD NNS | NNP PP | CD JJ NNS **| DT NNPS**

This change parsed the sentence as follows:

['the', 'Britons', 'migrate', 'south', 'frequently', '.']

(START

(S1

(NP (DT the) (NNPS Britons))

(VP (VB migrate) (RB south) (RB frequently))

(Eos .)))

#### Description

This sentence is parsed into a NP (noun phrase) consisting of a DT (determiner) “the” followed by a NNPS (plural proper noun) “Brightons” followed by a VP (verb phrase) which is made of up a VB (base verb) “migrate” followed by two RBs (adverbs) in succession “south” and “frequently” and then an Eos (End of sentence) “.”.

### Sentence 12

Arthur and Guinevere ride frequently near the castle .

In reviewing this sentence, we see that we have a proper noun “Arthur” followed by a coordinating conjunction “and” and then another proper noun “Guinevere”. This grammar needs to be added as well as the following verb phrase “ride frequently near the castle” which is a 3rd person plural verb “ride” followed by an advert and then a preposition phrase.

The following was changed to properly handle this sentence:

# Verb phrases

VP -> VBZ NP | VBZ NP PP | VBD VBN NP | VBP PP | MD VB VBN | VBD VBN VBG PP | MD VB VBN VBG NP | VB RB RB **| VBP RB PP**

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO | TO NP | DT JJS NNS | CD NNS | NNP PP | CD JJ NNS |DT NNPS **| NNP NP | CC NNP**

This change parsed the sentence as follows:

['Arthur', 'and', 'Guinevere', 'ride', 'frequently', 'near', 'the', 'castle', '.']

(START

(S1

(NP (NNP Arthur) (NP (CC and) (NNP Guinevere)))

(VP

(VBP ride)

(RB frequently)

(PP (IN near) (NP (DT the) (NN castle))))

(Eos .)))

#### Description

This sentence was parsed as a NP (noun phrase) consisting of a single NNP (proper noun) “Arthur” followed by another NP (noun phrase) which is a coordinating conjunction (CC) “and” followed by another NNP (proper noun) “Guinevere”. This is followed by a VP (verb phrase) which consists of a VBP (3rd person plural verb) “ride” then an RB (adverb) “frequently” followed by a PP (preposition phrase) which is an IN (preposition) “near” followed by another noun phrase which consists of a DT (determiner) “the” followed by a NN (noun) “castle” and an Eos (end of sentence) “.”.

### Sentence 13

he suggests to grow fruit at home .

In reviewing this sentence, we see that we have a verb phrase which is a singular 3rd person verb “suggests” followed by a TO and then a base verb and a noun phrase. This combination is not part of our existing grammar, to it needs to be added for parsing. We also have a noun phrase that consist solely of a personal pronoun. So, the following was changed to properly handle this sentence:

# Verb phrases

VP -> VBZ NP | VBZ NP PP | VBD VBN NP | | VBP PP | MD VB VBN | VBD VBN VBG PP | MD VB VBN VBG NP | VB RB RB | VBP RB PP **| VBZ2 TO VB NP**

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO | TO NP | DT JJS NNS | CD NNS | NNP PP | CD JJ NNS | DT NNPS | NNP NP | CC NNP **| PRP**

##### Sentence 13

This change parsed the sentence as follows:

(Eos .)))

['he', 'suggests', 'to', 'grow', 'fruit', 'at', 'home', '.']

(START

(S1

(NP (PRP he))

(VP

(VBZ2 suggests)

(TO to)

(VB grow)

(NP (NN fruit) (PP (IN at) (NP (NN home)))))

(Eos .)))

#### Description

This sentence was parsed as an NP (noun phrase) which is a PRP (personal pronoun) “he” followed by a VP (verb phrase) which consists of a VBZ2 (singular 3rd person verb) “suggests” followed by a TO (to) “to” and then a VB (base verb) “grow” and an NP (noun phrase) which consist of a NN (noun) “fruit” followed by a PP (preposition phrase) which is an IN (preposition) “at” then an NP (noun phrase) which is a single NN (noun) “home” and then an Eos (end of sentence) “.”.

### Sentence 14

riding to Camelot is not hard .

In reviewing this sentence, we have a special case where the first noun phrase consists of a present participle verb (VBG) “riding” followed by a noun phrase which consists of a TO (to) followed by a proper noun for a place “Camelot” . This is one situation where we can put this combination of grammar (VBG NP) in as a noun phrase (NP). The rest of the sentence is a VP starting with a VBZ (3rd person singular verb) “is” a NOT (not) “not” and then an adjective. These combinations were also not originally covered with our existing grammar.

The following was changed to properly handle this sentence:

# Verb phrases

VP -> VBZ NP | VBZ NP PP | VBD VBN NP | VBP PP | MD VB VBN | VBD VBN VBG PP | MD VB VBN VBG NP | VB RB RB | VBP RB PP | VBZ2 TO VB NP **| VBZ NOT JJ**

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO | TO NP | DT JJS NNS | CD NNS | NNP PP | CD JJ NNS | DT NNPS | NNP NP | CC NNP | PRP **|** **VBG NP | TO NNO**

This change parsed the sentence as follows:

['riding', 'to', 'Camelot', 'is', 'not', 'hard', '.']

(START

(S1

(NP (VBG riding) (NP (TO to) (NNO Camelot)))

(VP (VBZ is) (NOT not) (JJ hard))

(Eos .)))

#### Description

As mentioned in the reasoning for adding the grammar, this sentence is parsed as a NP (noun phrase) that starts with a VBG (vert past participle) “riding” followed by a NP (noun phrase) which consist of a TO (to) “to” and then a proper noun location/place (NNO) “Camelot” followed by a VP (verb phrase) which consist of a VBZ (3rd person singular verb) followed by NOT (not) “not” and then a JJ (adjective) “hard” ending with an Eos (end of sentence) “.”.

### Sentence 15

do coconuts speak ?

In reviewing this sentence, we have NP (noun phrase) which consists of DO (the word “do”) followed by a NNS (plural noun) and a VP which is a VB (base verb). This is the situation where we need to add a single VB to the verb phrases to handle “speak” and then a DO NNS to parse the “do coconuts” properly. We also need to add a sentence structure which is a verb phrase followed by a noun phrase and then an end of sentence.

The following was changed to properly handle this sentence:

# Sentences

S1 -> NP VP Eos **| VP NP Eos**

# Verb phrases

VP -> VBZ NP | VBZ NP PP | VBD VBN NP | VBP PP | MD VB VBN | VBD VBN VBG PP | MD VB VBN VBG NP | VB RB RB | VBP RB PP | VBZ2 TO VB NP | VBZ NOT JJ **| VB**

# Noun phrases

NP -> DT NP | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO | TO NP | DT JJS NNS | CD NNS | NNP PP | CD JJ NNS | DT NNPS | NNP NP | CC NNP | PRP | VBG NP | TO NNO **| DO NNS**

This change parsed the sentence as follows:

['do', 'coconuts', 'speak', '?']

(START (S1 (NP (DO do) (NNS coconuts)) (VP (VB speak)) (Eos ?)))

#### Description

These changes properly parse this sentence as a NP (noun phrase) consisting of DO “do” followed by a NNS (plural noun) “coconuts” and then a VP (verb phrase) which consists of a VB (base form verb) “speak” and then an Eos (end of sentence) “?”.

### Sentence 16

why does England have a king ?

In reviewing this sentence, we have a noun phrase that starts a WRB (a Wh-adverb) “why” followed by a noun phrase that uses DO (“does” ) followed by a proper noun location “England” and then another verb phrase which is “have a king” a base verb (VB) followed by a noun phrase which consists of a DT (determiner) and an NN (noun) “king”. We need to add something to handle the Wh-adverb as well as the “does England” and the “have” followed by a noun phrase:

# Verb phrases

VP -> VBZ NP | VBZ NP PP | VBD VBN NP | VBP PP | MD VB VBN | VBD VBN VBG PP | MD VB VBN VBG NP | VB RB RB | VBP RB PP | VBZ2 TO VB NP | VBZ NOT JJ | VB **| VB NP**

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO | TO NP | DT JJS NNS | CD NNS | NNP PP | CD JJ NNS | DT NNPS | NNP NP | CC NNP | PRP | VBG NP | TO NNO | DO NNS **| WRB NP | DO NNO**

This change parsed the sentence as follows:

['why', 'does', 'England', 'have', 'a', 'king', '?']

(START

(S1

(NP (WRB why) (NP (DO does) (NNO England)))

(VP (VB have) (NP (DT a) (NN king)))

(Eos ?)))

#### Description

This sentence is parsed as an NP (noun phrase) which consists of a WRB (Wh-adverb) “why” followed by a DO (do) “does” which is then followed by an NNO (proper noun object/location) “England”. This phrase is followed by a VP (verb phrase) which is a VB (base verb) “have” followed by an NP (noun phrase) which is a DT (determiner) “a” and then an NN (noun) “king” ending with an Eos (End of sentence) “?”.

### Challenge Sentence 1

what horse does Arthur ride ?

In reviewing this sentence, we Wh-determiner “what” followed by a base noun “horse” and then a noun phrase with a DO followed by a proper noun and then a verb phrase of a single base verb “ride”.

First, the following was changed to properly handle this sentence.

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO | TO NP | DT JJS NNS | CD NNS | NNP PP | CD JJ NNS | DT NNPS | NNP NP | CC NNP | PRP | VBG NP | TO NNO | DO NNS | WRB NP | DO NNO **| WDT NP | NN NP | DO NNP**

Although the sentence was parsed correctly, I modified some of the noun phrases that start with a DO ( DO NNS, DO NNO, etc) and set up a DO NP and then added noun phrases that represent the individual nouns NNS, NNO, etc.

Original Noun phrases:

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO | TO NP | DT JJS NNS | CD NNS | NNP PP | CD JJ NNS | DT NNPS | NNP NP | CC NNP | PRP | VBG NP | TO NNO **| DO NNS | WRB NP | DO NNO | WDT NP | NN NP | DO NNP**

Was modified to this:

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO | TO NP | DT JJS NNS | CD NNS | NNP PP | CD JJ NNS | DT NNPS | NNP NP | CC NNP | PRP | VBG NP | TO NNO **| DO NP | NNO | NNS | WRB NP | WDT NP | NN NP**

These changes parsed the sentence as follows:

['what', 'horse', 'does', 'Arthur', 'ride', '?']

(START

(S1

(NP (WDT what) (NP (NN horse) (NP (DO does) (NNP Arthur))))

(VP (VB ride))

(Eos ?)))

I also confirmed that all previous sentences were still parsed correctly with this modification to the noun phrases grammar.

#### Description

These changes properly parsed this first challenge sentence as a NP (noun phrase) consisting of WDT (Wh-determiner) “what” then another NP (noun phrase) which is a single NN (noun) “horse” and then a final NP (noun phrase) which consists of a DO “does” followed by a NNP (proper noun) “Arthur” ending with a VP (verb phrase) of a VB (single base verb) “ride” and then an Eos (end of sentence) “?”.

### Challenge Sentence 2

who does Arthur suggest she carry ?

In reviewing this sentence, we Wh-pronoun “who” (WP) followed by a DO noun phrase and then a verb phrase with a noun phrase of a personal pronoun followed by a verb phrase of a base verb “carry”. These are not currently handled by our grammar, so the following was changed to properly handle this sentence.

# Noun phrases

NP -> DT NN | NNP | NN PP | NN | DT JJ NN | DT NNO | DT JJ NNO | TO NP | DT JJS NNS | CD NNS | NNP PP | CD JJ NNS | DT NNPS | NNP NP | CC NNP | PRP | VBG NP | TO NNO | DO NNS | WRB NP | DO NNO | WDT NP | NN NP | DO NNP **| WP NP | PRP VP**

These changes parsed the sentence as follows:

['who', 'does', 'Arthur', 'suggest', 'she', 'carry', '?']

(START

(S1

(NP (WP who) (NP (DO does) (NP (NNP Arthur))))

(VP (VB suggest) (NP (PRP she) (VP (VB carry))))

(Eos ?)))

#### Description

These changes properly parsed this second challenge sentence as a NP (noun phrase) consisting of WP (Wh-pronoun) “who” and another NP (noun phrase) which consists of a DO “does” followed by a NP (noun phrase) of a single NNP (proper noun) “Arthur” ending with a VP (verb phrase) of a VB (single base verb) “suggest” then a NP (noun phrase) of a PRP (personal pronoun) “she”, a VB (base verb) “carry” and then an Eos (end of sentence) “?”.

# Part 2 – Exemplar Sentences

## Sentence That Cannot be Parsed (Overfitting)

another plodding servant ought to master the simplest horse .

#### Description

This sentence contains all the same words in our grammar but they are used in a different manner than our rules cover. We have a starting NP (noun phrase) “another plodding servant” which is a DT (determiner) “another” followed by a JJ (adjective) and then a NN (noun) “servant” (a covered sequence of *DT JJ NN*) and then a MD (modal) “ought” and a PP (preposition phrase) which is a “to master the simplest horse”. However, we have used the word ‘master’ as a verb and not a noun as it is outlined in our grammar which is why this cannot be parsed correctly. In addition, our PP phrase is only a IN (preposition) followed by a NP (noun phrase) and this is a PP followed by a VP (VB “master”) and then a NP “the simplest horse” and an Eos (End of sentence) “.”.

## Gibbersh Sentence that Parses (Overgeneralization)

Gibberish sentence that parses

which castle drinks the weariest mountains.

This sentence contains all the same words in our grammar in an order that can be parsed with our original grammar, but the words make no sense in the order provided. However, it is parsed as follows:

['which', 'castle', 'drinks', 'the', 'weariest', 'mountains', '?']

(START

(S1

(NP (WDT which) (NP (NN castle)))

(VP (VBZ drinks) (NP (DT the) (JJS weariest) (NNS mountains)))

(Eos ?)))

#### Description

This sentence is parsed as a NP (noun phrase) which is a WDT (Wh-determiner) “which” followed by a NP (noun phrase) which consists of a single NN (noun) “castle”. This is followed by a VP (verb phrase) of a VBZ (3rd person singular verb) “drinks” followed by a NP (noun phrase) that consists of a DT (determiner) “the” and then a JJS (superlative adjective) “weariest” and then an NNS (plural noun) “mountains” and then an Eos (End of sentence) “?”.

# Conclusion

This exercise was challenging, but enjoyable. It was interesting trying to find the best way to handle the tagging and parsing.