Probabilistic Context Free Grammar

Siri Chandana S(829955672)

Part 1- Writing grammar rules

For each sentence in the ‘dev.sen’ file, the grammar rules were changes in the ‘S1.gr’ file and the ‘Vocab.gr’ file. After making the changes to the rules the parse tree was obtained by parsing through the ‘S1.gr’ grammar rules rather than ‘S2.gr' rules. The approach used for finding the rules and generating the parse tree is that first extract all the sentences from the ‘dev.sen' file
Sentence: Arthur rides the plodding horse near the castle. For each sentence, assign a POS tag to the word which is MISC and then add rules if the current rules donot generate parse trees. So the following are the changes made to the ‘Vocab.gr’ and ‘S1.gr’ for each sentence-

- **Sentence 1**- Arthur rides the plodding horse near the castle .

‘Vocab.gr’ file- All the words in the sentence are tagged except for ‘plodding’. This word is an adjective and since the adjective terminal word doesn’t exist, ‘Adj’ is used to define adjectives. So, ‘plodding’ is marked as ‘Adj’

‘S1.gr’ file - Since, there are no rules for adjective phrases. The following rule was added to parse noun phrases which have adjectives in them-

1. NP Det Adj Nbar- this grammar rule helps in parsing the “the plodding horse” part of the sentence.

Parse Tree- The following is the tree obtained for the grammar in ‘S1.gr’ file-

```
(START
   (S1
      (@NP_VP
         (NP
            (Proper Arthur))
         (VP
            (VerbT rides)
            (NP
               (@Det_Adj
                  (Det the)
                  (Adj plodding))
               (Nbar
                  (Nbar
                     (Noun horse))
                  (PP
                     (Prep near)
                     (NP
```

```
                    (Det the)
                    (Nbar
                       (Noun castle))))))))) .))
```

- **Sentence 2**- the Holy Grail is a chalice .

'Vocab.gr' file- All the words in the sentence are tagged except for 'Holy Grail'. This word is an proper noun and all proper nouns have terminal symbol 'Proper'. So, 'Holy Grail' is marked as 'Proper'.

'S1.gr' file - The following rule was added to parse noun phrases which have proper nouns in them rather than noun phrases which are only proper nouns-

1. NP Det Proper- with the help of this grammar rule the proper noun "Holy Grail" can be parsed by the parser as it initially couldn't handle these phrases.

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@NP-VP
         (NP
            (Det the)
            (Proper Holy) Grail))
         (VP
            (VerbT is)
            (NP
               (Det a)
               (Nbar
                  (Noun chalice))))) .))
```

- **Sentence 3**- the sensational Holy Grail is a sacred chalice

'Vocab.gr' file- All the words in the sentence are tagged except for 'sensational' and 'sacred'. These words is are both adjectives. So, 'sensational' and 'sacred' is marked as 'Adj'.

'S1.gr' file - The following rule was added to parse noun phrases which have adjectives that are followed by proper nouns-

1. NP Det Adj NP- With the help of this rule the first of the sentence " the sensational Holy Grail" can be parsed by the parser.

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@NP-VP
         (NP
            (@Det-Adj
```

```
            (Det the)
            (Adj sensational))
         (NP
            (Proper Holy) Grail)))
      (VP
         (VerbT is)
         (NP
            (@Det-Adj
               (Det a)
               (Adj sacred))
            (Nbar
               (Noun chalice))))) .))
```

- **Sentence 4**- every coconut was carried to the hottest mountains

'Vocab.gr' file- The word 'to' in this sentence is marked as 'To'(creating a new non-terminal). The words 'was' and 'carried' are past tense verbs so, they are marked as VerbPT (creating a new non-terminal). The word 'hottest' is an adjective and marked as 'Adj' and the word mountains is marked as 'NP'.

'S1.gr' file - The following rules are added to parse different types on verb phrases-

1. VP VPZ -This rule is to handle different kinds of verb phrases like "was".
2. VP VPZ VP- this rule will help us handle different combinations of verb phrases like in this sentence "was" is followed by "carried"
3. VP VP To NP- This mainly deals with the To phrases that occur before noun phrases and after verb phrases like "carried to the".
4. VPZ VerbT- This is rule added the terminal symbol VerbT to the grammar.
5. VPZ VerbPT- This is rule added the terminal symbol VerbPT that is past tense verbs to the grammar.

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@NP-VP
         (NP
            (Det every)
            (Nbar
               (Noun coconut)))
         (VP
            (VPZ
               (VerbPT was))
            (VP
               (@VP-To
                  (VP
```

```
            (VPZ
               (VerbPT carried)))
         (To to))
      (NP
         (@Det-Adj
            (Det the)
            (Adj hottest))
         (NP mountains))))) .))
```

- **Sentence 5**- sixty strangers are at the Round Table

'Vocab.gr' file- The word 'sixty' is a number but in this sentence it is being used an adjective so, it is marked as 'Adj' . 'Round Table' is marked as 'Proper' as it is a proper noun and 'are' is a verb which is marked as 'VerbT'(general verb).

'S1.gr' file - The following rules are added to parse different types on verb phrases and the noun phrases-

1. VP VerbT PP- This rule will help parse prepositions which follow verbs like "are at" in this sentence.
2. NP Adj NP- This rule will help parse noun phrases that follow adjectives like "sixty strangers" in this sentence.

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@NP-VP
         (NP
            (Adj sixty)
            (NP strangers))
         (VP
            (VerbT are)
            (PP
               (Prep at)
               (NP
                  (Det the)
                  (Proper Round) Table))))) .))
```

- **Sentence 6**- Sir Lancelot might have spoken

'Vocab.gr' file- The word 'might' and 'spoken' are marked as 'VerbMT'(creating new terminal symbol for modals" and 'VerbT'(general verbs) respectively. However, 'have' is a base form of the verb so it is marked as 'VerbBF'.

'S1.gr' file - The following rules are added to parse different types on verb phrases-

1. VPZ VerbMT VerbBF- this rule is to parse different types of verbs that is when a modal is followed by a base form verb like "might have".

Parse Tree- The following is the tree obtained for the grammar in ‘S1.gr’ file-

```
(START
   (S1
      (@NP-VP
         (NP
            (Proper Sir) Lancelot))
         (VP
            (VPZ
               (VerbMT might)
               (VerbBF have))
            (VP
               (VPZ
                  (VerbT spoken))))) .))
```

- **Sentence 7**- Guinevere had been riding with Patsy for five weary nights

‘Vocab.gr’ file- The word ‘five’ is a number but in this case it is used as a noun phrase so it is marked as ‘NP’ and the word ‘weary’ is an adjective so it is marked as ‘Adj’. Finally, the word ‘nights’ is used as a noun phrase so it is tagged as ‘NP’.

‘S1.gr’ file - The following rules are added to parse different types on verb phrases and noun phrases-

1. VP VP PP NP- this rule is used for sections of the sentences like “riding with patsy”.

Parse Tree- The following is the tree obtained for the grammar in ‘S1.gr’ file-

```
(START
   (S1
      (@NP-VP
         (NP
            (Proper Guinevere))
         (VP
            (VPZ
               (VerbT had))
            (VP
               (VPZ
                  (VerbT been))
               (VP
                  (@VP-PP
                     (VP
                        (VerbT riding)
                        (PP
                           (Prep with)
                           (NP
                              (Proper Patsy))))
                     (PP
                        (Prep for)
                        (NP five)))
```

```
                  (NP
                     (Adj weary)
                     (NP nights)))))) .))
```

• **Sentence 8**- Sir Bedevere might have been suggesting this quest

'Vocab.gr' file- The word 'suggesting' is tagged as 'VerbT' as it is a verb(general verb).

'S1.gr' file - No rules are added to the grammar, the grammar is sufficient to parse this sentence.

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@NP_VP
         (NP
            (Proper Sir) Bedevere))
         (VP
            (VPZ
               (VerbT might)
               (VerbBF have))
            (VP
               (VPZ
                  (VerbT been))
               (VP
                  (VerbT suggesting)
                  (NP
                     (Det this)
                     (Nbar
                        (Noun quest))))))) .))
```

• **Sentence 9**- the Britons migrate south frequently

'Vocab.gr' file- The word 'migrate' is tagged as 'VerbT' as it is a verb(general verb). The word 'Britons' is a proper noun and hence it is tagged as 'Proper'. Both the words south and frequently are adverbs and they are tagged as 'Adv' (creating terminal).

'S1.gr' file - ADVP is a new non-terminal that has been added to the grammar to handle the adverb phrase. The following rules are added to the grammar to parse verb phrases that consist of adverb phrases-

1. VPZ VerbBF- sentences that have verbs like "migrate"
2. VPZ VPZ ADVP- this used to parse sentences that have adverbs following verb phrases like "migrate south".
3. ADVP Adv- this is for adverbs like "south" which exist in a sentence.

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@NP-VP
         (NP
            (Det the)
            (Proper Britons))
         (VP
            (VPZ
               (VPZ
                  (VPZ
                     (VerbBF migrate))
                  (ADVP
                     (Adv south)))
               (ADVP
                  (Adv frequently))))) .))
```

- **Sentence 10**- Arthur and Guinevere ride frequently near the castle

'Vocab.gr' file-All the words have been tagged except 'and'. As 'and' is a conjugation, it is tagged as 'CC'.

'S1.gr' file - The following rules are added to the grammar to parse noun phrases that consist of conjugations and also the adverb phrases are modified-

1. NP Proper CC Proper- this rule is to handle sections of sentences where proper nouns are separated with conjunctions like "Arthur and Guinevere".
2. ADVP Adv PP- this rule is for prepositional phrases that follow adverbs like "frequently near the castle".

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@NP-VP
         (NP
            (@Proper-CC
               (Proper Arthur)
               (CC and))
            (Proper Guinevere))
         (VP
            (VPZ
               (VPZ
                  (VerbBF ride))
               (ADVP
                  (Adv frequently)
                  (PP
                     (Prep near)
```

```
            (NP
               (Det the)
               (Nbar
                  (Noun castle))))))))) .))
```

- **Sentence 11**- he suggests to grow fruit at home

'Vocab.gr' file- 'he' is a personal pronoun and it is tagged as 'PRP'(creating new terminal). The word 'suggests' is a verb and tagged as 'VerbT' and the word 'grow' is a verb's base form so it is tagged as 'VerbBF'.

'S1.gr' file - The following rules are added to the grammar to parse verb phrases and nouns present in noun phrases-

1. VP VPZ NP
2. VP VP To VP
3. NP Nbar

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@NP_VP
         (NP
            (PRP he))
         (VP
            (@VP_To
               (VP
                  (VPZ
                     (VerbT suggests)))
               (To to))
            (VP
               (VPZ
                  (VerbBF grow))
               (NP
                  (Nbar
                     (Nbar
                        (Noun fruit))
                     (PP
                        (Prep at)
                        (NP
                           (Nbar
                              (Noun home))))))))) .))
```

- **Sentence 12**- riding to Camelot is not hard .

'Vocab.gr' file- The word 'Camelot' is a proper noun and is tagged as 'Proper'. 'Not' is tagged with 'Not' by creating a new terminal and the word 'hard' is an adjective in this sentence. So, it is tagged as 'Adj'. In this sentence the word "riding" is a noun and hence, it is tagged as "Noun".

'S1.gr' file - The following rules are added to the grammar to handle phrases with 'Not' in them-

1. NP NP to NP - this rule parses noun phrases that have to in them like "riding to Camelot".
2. VP VerbT Not Adj- this rule is to parse parts of sentences like "is not hard" which are verb phrases.

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@NP_VP
         (NP
            (@NP_To
               (NP
                  (Nbar
                     (Noun riding)))
               (To to))
            (NP
               (Proper Camelot)))
         (VP
            (@VerbT_Not
               (VerbT is)
               (Not not))
            (Adj hard))) .))
```

- **Sentence 13**- do coconuts speak ?

'Vocab.gr' file- The word 'do' is tagged as 'Vdo' by creating a new terminal word and the word 'coconuts' is tagged as 'noun phrase'.

'S1.gr' file - The start symbol was modified to handle sentences which are questions. The following rules are added to the grammar to handle phrases which have 'do' in them-

1. S1 VBP NP VP ? - rule has been added to S1 to handle questions like these ending with "?".
2. VBP Vdo- This is to handle the starting part of the sentence ending in question mark like "do".

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@@VBP_NP_VP
         (@VBP_NP
            (VBP
               (Vdo do))
            (NP coconuts))
         (VP
            (VPZ
               (VerbT speak)))) ?))
```

- **Sentence 14**- why does England have a king ?

'Vocab.gr' file- The word 'why' is tagged as 'Adv' as it is an adverb and the word 'does' is tagged as 'Vdo' as it is a variation of the word 'do'. Also, the word 'England' is tagged as 'Proper' as it is a proper noun.

'S1.gr' file - The following rules are added to the grammar to handle adverb phrases which have 'do' in them-

1. VBP Adv Vdo- this is to handle adverbs which are followed by verbs like "why does" in this sentence.

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@@VBP_NP_VP
         (@VBP_NP
            (VBP
               (Adv why)
               (Vdo does))
            (NP
               (Proper England)))
         (VP
            (VPZ
               (VerbBF have))
            (NP
               (Det a)
               (Nbar
                  (Noun king))))) ?))
```

Sentences from the Challenge sentences

Now, from the 'challenge.sentences.txt', the below two sentences are chosen to be parsed by our grammar in the 'S1.gr' file.

Sentence 1- what horse does Arthur ride ?

'Vocab.gr' file- The word 'what' is tagged as 'Whp' as these are the determiners(creating new terminal symbol).

'S1.gr' file - The following rules are added to the grammar to handle determiner phrases in them-

1. VBP Whp NP VBP- this rule is to parse sentences that start with determiners like "what" in this sentence.

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@@VBP_NP_VP
         (@VBP_NP
            (VBP
               (@Whp_NP
                  (Whp what)
                  (NP
                     (Nbar
                        (Noun horse))))
               (VBP
                  (Vdo does)))
            (NP
               (Proper Arthur)))
         (VP
            (VPZ
               (VerbBF ride)))) ?))
```

Sentence 2- who does Arthur suggest she carry ?

'Vocab.gr' file- The word 'who' is tagged as 'Whp' as these are the determiners(creating new terminal symbol). The verbs 'suggest' and 'carry' are base forms so they are tagged as 'VerbBF' and she in this case is a personal pronoun so it is tagged as 'PRP'.

'S1.gr' file - The following rules are added to the grammar to handle determiner phrases which have 'do' in them-

1. VBP Whp Vdo - this rule is for sentences that start with "who does".
2. VP VP NP VP - this rule is to handle sentences which are having verb phrases following noun phrases and before as well like "suggest she carry".

Parse Tree- The following is the tree obtained for the grammar in 'S1.gr' file-

```
(START
   (S1
      (@@VBP_NP_VP
         (@VBP_NP
            (VBP
               (Whp who)
               (Vdo does))
            (NP
               (Proper Arthur)))
         (VP
            (@VP_NP
               (VP
                  (VPZ
                     (VerbBF suggest)))
               (NP
                  (PRP she)))
```

```
            (VP
               (VPZ
                  (VerbBF carry))))) ?))
```

Part 2 -Exemplar Sentences

1) English Sentence-

- the castle was Arthur 's home . - This sentence is an English sentence which has possessive words like “Arthur’s” which is not handled by our grammar. Hence, this sentence will be parsed by S2 grammar. The parse tree is as follows-

```
(START
   (S2
      (_NP
         (NP
            (Det the)
            (Nbar
               (Noun castle)))
         (_VerbPT
            (VerbPT was)
            (_Proper
               (Proper Arthur)
               (_Misc
                  (Misc 's)
                  (_Noun
                     (Noun home)
                     (_Misc
                        (Misc .)))))))))
```

- neither Patsy nor Arthur spoke again . - This sentence has coordinating conjunctions like “neither” and “nor” which are not handled by our grammar. Hence, this sentence will be parsed by S2 grammar. The parse tree is as follows-

```
(START
   (S2
      (_Misc
         (Misc neither)
         (_Proper
            (Proper Patsy)
            (_Misc
               (Misc nor)
               (_Proper
                  (Proper Arthur)
                  (_Misc
                     (Misc spoke)
                     (_Misc
                        (Misc again)
                        (_Misc
                           (Misc .)))))))))))
```

2) Non-english sentence-

- do Arthur speak horse ? - This sentence is semantically wrong as horse is not a language and it cannot be spoken. This sentence is also syntactically wrong as do should be present before a plural but here Arthur is a singular word and singular words should have does before them and not do. This happens because in the grammar both do and does have the same terminal symbol and this aspect of the english sentences is not considered by the grammar. The parse tree that is obtained for this sentence is a follows-

```
(START
   (S1
      (@@VBP_NP_VP
         (@VBP_NP
            (VBP
               (Vdo do))
            (NP
               (Proper Arthur)))
         (VP
            (VerbT speak)
            (NP
               (Nbar
                  (Noun horse))))) ?))
```

The rules that are used to parse this sentence are-

1. S1 VBP NP VP ?
2. VBP Vdo
3. NP Proper
4. VP VerbT NP
5. NP Nbar
6. Nbar Noun

Conclusion-

The grammar present in “S1.gr" has successfully parsed all the sentences in the “dev.sen” file.