

Homework 2

Brandon Bluemner

Question 1

Suppose a grammar allows conjoined constructions such as

$S \rightarrow S \text{ and } S$ $S \rightarrow S \text{ while } S$

$S \rightarrow S \text{ or } S$ $S \rightarrow S \text{ when } S$

$S \rightarrow S \text{ but } S$

Provide two correct parse trees for the following sentence, using nested, bracketed notation. (Note that this is another one that the MS SPLAT demo parser gets totally wrong.)

John slept while Mary ate and the dog barked

1)

Exploded view

```

[S
  [S
    [NP
      [NNP John]
      [VP
        [VBD slept]
      ]
    ]
  ]
]
while
[S
  [S
    [NP
      [NNP Mary]
    ]
    [VP
      [VBD ate]
    ]
  ]
]
and
[S
  [NP
    [DT the]
    [NN dog]]
  [VP
    [VBN barked]
  ]
]
]
]

```

Compressed view

```

[S [S [NP [NNP John][VP [VBD slept]]]] while [S [S [NP [NNP Mary]][VP [VBD ate]]] and [S
[NP [DT the][NN dog]] [VP [VBN barked]]]]]

```

2)

Exploded view

```

[S
  [S
    [S
      [NP
        [NNP John]
      ]
      [VP
        [VBD slept]
      ]
    ]
    while
    [S
      [NP
        [NNP Mary]
      ]
      [VP
        [VBD ate]
      ]
    ]
  ]
  and
  [S
    [NP
      [DT the]
      [NN dog]
    ]
    [VP
      [VBN barked]
    ]
  ]
]
```

Compressed view

```

[S [S [S [NP [NNP John]][VP [VBD slept]]] while [S [NP [NNP Mary]][VP [VBD ate]]] ] and [S
[NP [DT the][NN dog]] [VP [VBN barked]]]]
```

Question 2

Consider the following long literary sentences and for each part provide the following:

- i. Give a set of all the additional S-level rules that you would need including any additional conjunctions or other function words.
- ii. Show each passage with brackets around all the complete sentences and noun phrases (possibly nested) based on the combined S grammar (from part 2i and question 1) and your understanding of noun phrases (e.g. from homework 1).

a) E.B. White, "Stuart Little." 107 words.

"In the loveliest town of all, where the houses were white and high and the elms trees were green and higher than the houses, where the front yards were wide and pleasant and the back yards were bushy and worth finding out about, where the streets sloped down to the stream and the stream flowed quietly under the bridge, where the lawns ended in orchards and the orchards ended in fields and the fields ended in pastures and the pastures climbed the hill and disappeared over the top toward the wonderful wide sky, in this loveliest of all towns Stuart stopped to get a drink of sarsaparilla."

Let the S-level be define by the following grammer:

S -> NP VP | PP PUNCT SBAR | S CC S | S PUNCT SBAR | S PUNCT

NP -> DT JJ NN | DT NN NNS | DT JJ NNS | DT JJS NN| DT ADJP NN | PP NNP | DT NN | DT NNS | NP PP | NN

PP -> IN NP|IN PP| IN DT | PP PP | IN ADJP | TO NP | IN NNS | IN NN | DT JJ

ADJP -> JJ CC JJ | JJ CC JJR | DT JJS | PP ADJP | ADJP PP | JJ JJ

SBAR -> WHADVP S | WHADVP

WHADVP -> WRB

VP -> VP CC VP | VBD RB PP | VBD ADJP VP| VBG RP RB | VBD ADJP | VBD NP | VP VP | VB NP | TO VP | VBD PP| VBN PP| VBDS

IN -> 'In'|'of'|'than'|'under'|'in'|'over'|'toward'

VBG -> 'finding'

WRB -> 'where'

JJS -> 'loveliest'

VBN -> 'ended'

VB -> 'get'

RP -> 'out'

JJR -> 'higher'

CC -> 'and'

RB -> 'about'|'down'|'quietly'

JJ -> 'white'|'high'|'green'|'wide'|'pleasant'|'back'|'bushy'|'worth'|'top'|'wonderful'

NNP -> 'Stuart'

TO -> 'to'

DT -> 'the'|'all'|'this'|'a'

VBD -> 'were'|'sloped'|'flowed'|'ended'|'climbed'|'disappeared'|'stopped'

PUNCT -> ','|'.'

NN -> 'town'|'elms'|'front'|'stream'|'bridge'|'lawns'|'hill'|'sky'|'drink'|'sarsaparilla'

NNS -> 'houses'|'trees'|'yards'|'streets'|'orchards'|'fields'|'pastures'|'towns'

Which yields the following:

[S [PP [IN In]
 [NP [NP [DT the][JJS loveliest][NN town]]
 [PP [IN of][DT all]]]]
 [PUNCT ,]
 [SBAR
 [WHADVP[WRB where]]
 [S [NP [DT the][NNS houses]]
 [VP [VBD were][UCP[JJ white][CC and][JJ high]]]]]
 [CC and]
 [S [NP [DT the]
 [NN elms]
 [NNS trees]
]
 [VP [VBD were]
 [ADJP[ADJP[JJ green][CC and][JJR higher]]
 [PP[IN than][NP[DT the][NNS houses]]]]]
 [PUNCT ,]
 [SBAR
 [WRB where]
 [S [NP [DT the][NN front][NNS years]]
 [VP [VBD were][ADJP[JJ wide][CC and][JJ pleasant]]]]
 [CC and]
 [S [NP [DT the][JJ back][NNS yards]]
 [VP [VBD were]
 [ADJP [JJ worth][CC and][JJ worth]]
 [VP [VGB finding][PRT[RP out]][RB about]]]
 [PUNCT ,]
 [SBAR
 [WHADVP [WRB where]]
 [S [NP [DT the][NNS streets]]
 [VP [VBD sloped][PRT[RB down]][PP[TO to][NP[DT the][NN stream]]]]]
 [CC and]
 [S [NP [DT the][NN stream]]
 [VP [VBD flowed][PRT[RB quietly]][PP [IN under][NP [DT the][NN bridge]]]]
 [PUNCT ,]
 [SBAR [WHADVP [WRB where]]
 [S [NP[DT the][NN lawns]]
 [VP[VBN ended][PP [IN in][NNS orchards]]]]]
 [CC and]
 [S [NP [DT the][NNS orchards]]
 [VP [VBD ended][PP[IN in][NNS fields]]]
 [CC and]
 [S [NP [DT the][NNS fields]]
 [VP [VBD ended][PP[IN in][NNS pastures]]]
 [CC and]
 [S [NP[DT the][NNS pastures]]
 [VP [VP[VBD climbed][NP [DT the][NN hill]]]
 [CC and]
 [VP[VBD disappeared]

```
[PP [IN over]
    [PP[PP[DT the]
        [JJ top]]
        [PP[NP [DT the]
            [ADJP[JJ wonderful][JJ wide]]
            [NN sky]]]]]]]]
[PUNCT ,]
[S
    [NP
        [PP      [IN in][PP [ADJP[DT this][JJS loveliest]]]
                [PP[IN of][NP[DT all][NNS towns]]]]
        [NNP Stuart]]
    [VP [VBD stopped]
        [S [VP [TO to]
            [VP[VB get]
                [NP[NP[DT a][NN drink]]
                    [PP[IN of]
                        [NP[NN sarsaparilla]]]]]]]]]]]
```

Issues

This size of the copra made testing difficult using the `nltk.CFG.fromstring` as memory sizes exceed my system 16gb, due to some of the recursion in the grammer, so some opitimization is needed.

b) (GRAD ONLY) W.G. Sebald, "The Rings of Saturn." 107 words.

"All I know is that I stood spellbound in his high-ceilinged studio room, with its north-facing windows in front of the heavy mahogany bureau at which Michael said he no longer worked because the room was so cold, even in midsummer; and that, while we talked of the difficulty of heating old houses, a strange feeling came upon me, as if it were not he who had abandoned that place of work but I, as if the spectacles cases, letters and writing materials that had evidently lain untouched for months in the soft north light had once been my spectacle cases, my letters and my writing materials"

S -> NP VP | PP PUNCT SBAR | S CC S | S PUNCT SBAR | S PUNCT | SBAR PUNCT NNS CC S | VP | NP
ADVP VP

NP -> DT JJ NN | DT NN NNS | DT JJ NNS | DT JJS NN | DT ADJP NN | PP NNP | DT NN | DT NNS | NP PP |
NN | NP SBAR | NP PP SBAR | NP SBAR

PP -> IN NP | IN PP | IN DT | PP PP | IN ADJP | TO NP | IN NNS | IN NN | DT JJ | AVBP IN NP

ADJP -> JJ CC JJ | JJ CC JJR | DT JJS | PP ADJP | ADJP PP | JJ JJ | RB JJ

ADBP -> RB

SBAR -> WHADVP S | WHADVP | SBAR SBAR | S | IN SBAR

WHADVP -> WRB

VP -> VBD PP PUNCT SBAR PRP | VP CC VP | VBD RB PP | VBD ADJP VP | VBG RP RB | VBD ADJP | VBD
NP | VP VP | VB NP | TO VP | VBD
PP | VBN PP | VBDS | VBD AVBP VP |

```

[S
  [S
    [NP [DT ALL][SBAR [S [NP I][VBP now]]]]]
    [VP [VBZ is]
      [SBAR [IN that][S [
        [NP [PRP I]]
        [VP
          [VBD stood]
          [NP [NN speelbound]
            [PP [IN in]
              [NP [PRP$ his] [JJ high-eilinged][NN studio][NN room]]
            ]
          ]
        ]
        [PUNCT]
        [PP [IN with]
          [NP [PRP$ its]
            [JJ north-facing]
            [NNS windows]
          ]
        ]
        [NP [IN in]
          [NP front]
          [PP [IN of]
            [NP [DT the]
              [JJ heavy]
              [NN mahogany]
              [NN bureau]
            ]
          ]
          [SBAR [WHPP [IN at]]
            [WHNP [WDT which]]
            [S [NP [NNP Michael]]
              [VP [VBD said]
                [SBAR [S [NP [PRP he]]
                  [ADVP [RB no]
                    [RB longer]
                  ]
                ]
              ]
            ]
            [VP [VBN worked]
              [SBAR[IN because]
                [S [NP [DT the][NN room]]
                  [VP [ADJP [RB so][JJ col
                    [PUNCT ,]
                    [PP [ADVBR [RB even]]
                      [IN in]
                      [NP [NN midsumme
r]]]

```

```

]
]
]
]
[PUNCT ;]
]
]
]
]
]
]
]
]
]]]
]
]
[CC and]
[S
  [SBAR [SBAR [IN that]
    [PUNCT ,]
    [S [SBAR [IN while]
      [S [NP [PRP we]]
        [VP [VBD talked]
          [PP [IN of]
            [NP [NP[DT the][NN difficulty]]
              [PP [IN of]
                [VP[VGB heating][NP[JJ old][NNS houses]] ]
              ]
            ]
          ]
        ]
      ]
    ]
  ]
  [PUNCT ,]
  [NP [DT a][JJ strange][NN felling]]
  [VP [VBD came]
    [PP [IN upon] [NP [PRP me]]]
  ]
  [PUNCT ,]
  [SBAR [IN as][SBAR [IN if][S [NP [PRP it][VP[VBD were][RB not]]
]]]]
  [PRP he]
]
]
]
[SBAR [WP who]
  [S [VP [VBD had]
    [VP [VBN abandoned]
      [PP [IN that][NN place]]

```

]

Question 3

Consider the following examples of “garden path” (GP) sentences and some similar non-garden path sentences. (They are from the Wikipedia entry on Garden Path Sentences). Using the MSR SPLAT demo (<http://msrsplatdemo.cloudapp.net/>), compute the precision and recall of the GP versions based on the number of correct phrases when compared to the non-GP versions. (Show the unsimplified expressions that you used to compute the precision and recall.).

$$\text{Precision} = \frac{\text{Number of correct cunks given by system}}{\text{Total number of chunks given by system}}$$

$$\text{Recall} = \frac{\text{Number of correct cunks given by system}}{\text{Total number of actual chunks in the text}}$$

a) The horse raced past the barn fell. vs The horse that was raced past the barn fell.

Actual Chunker

`["[NP The horse] [VP raced] [PP past] [NP the barn] [VP fell] \n"]`

Total 5 chunks

Expected Chunker

`["[NP The horse] [NP that] [VP was] [VP raced] [ADJP past] [NP the barn] [VP fell] . \n"]`

Total 7 chunks

Correct Cunks

`[NP The horse] [VP raced] [NP the barn] [VP fell]`

Total 4

Assuming the numbers above are correct

$$\text{Precision} = \frac{4}{5} \approx 80\%$$

$$\text{Recall} = \frac{4}{7} \approx 57.143\%$$

b) The old man the boats vs The old sank the boats.

Actual Chunker

["[NP The old man] [NP the boats] \n"]

Total 2

Epected Chunker

["[NP The old] [VP sank] [NP the boats] \n"]

Total 3

Correct Cunks

Total chunks

$$Precision = \frac{1}{2} \approx 50\%$$

$$Recall = \frac{1}{3} \approx 33\%$$

c) The complex houses married and single soldiers and their families. vs The complex was sold to married and single soldiers and their families

Actual Chunker

["[NP The complex houses] [NP married and single soldiers] and [NP their families] . \n"]

Total 3

Epected Chunker

["[NP The complex] [VP was sold] [PP to] [NP married and single soldiers] and [NP their families] . \n"]

Total 5

Correct Cunks

Total chunks

$$Precision = \frac{2}{3} \approx 66.6\%$$

$$Recall = \frac{2}{5} \approx 40\%$$

Question 4 Programming Question (probabilistic phrase structure grammars and chart parsers)

a) How many productions are there in the treebank that start with 'S'? How many start with NP (or a variant)?

len 'S':772

len 'NP':1490

len 'NP or variant include WHNP*':3347

b) Describe what happens after changing the beam width and what that signifies in terms of the grammar

This is increasing the size of the queue, this means that we discard items in the queue is larger than the beam, by increasing the beam the queue has more the chart allowing for a larger "search" area, aka more information.

c) What was the smallest value you found for BIGGERNUM that finds the correct parse? Provide a table that shows different beam widths and the number of parses returned, where you include at least one beam value for each possible number.

Width	Parses Returned
1	0
1042	1
1109	2
1170	3

d) Create a new file, runGrammar2.py that defines a new grammar (a modification of the one provided), defines a new function that reads the sentences from an input file, and then runs the parser on the sentences from the file and prints the output to another file called hw2_output.txt. Your updated toy grammar, called grammar2, should add grammatical and lexical productions (or add new items on the RHS of existing rules) as needed to handle:

- imperative sentences such as "eat the salad",
- ditransitive sentences, such as "I read her the book" (but not *I read she the book")
- passive sentences, such as "The book was read to her" (but not *I read the book to she)

You should add a new category of pronoun, Pro_obj, for pronouns that can be indirect objects or objects of prepositions (such as him, her, them) but not subjects. Create a text file, hw2_input.txt with enough test sentences, each on a separate line, to demonstrate that your grammar works correctly, and then run your runGrammar2.py code.

