Welcome to the Statistical Methods of Language Technologyb SoSe21 course

Dr. Seid Muhie Yimam

- Email: yimam@informatik.uni-hamburg.de (mailto:yimam@informatik.uni-hamburg.de)
- Office: Informatikum, F-415

Dr. Özge Alaçam

- Email: alacam@informatik.uni-hamburg.de (mailto:alacam@informatik.uni-hamburg.de)
- Office: Informatikum, F-435

Topic of this week;

In this first practice class, we are going to focus on two main topics, which will be useful to complete the assignment;

- · CFGs and PCFG parsing
- Parseval

Deadline: 09/07 June

In class Exercises

Problem 7.1 CFGs and PCFG parsing

a) Convert this CFG to CNF. Capital letters are non-terminals and lowercase letters are terminals, S is the start symbol.

$$S \rightarrow A B C \mid S D$$

$$A \rightarrow D D \mid \epsilon$$

$$B \rightarrow F \mid C C C$$

$$C \rightarrow a c \mid a d \mid a e \mid \epsilon$$

$$D \rightarrow C d d \mid d e$$

$$F \rightarrow \epsilon$$

Context Free grammar to Chomsky normal form (CNF) conversion

···· , ···············

CONVERSION TO CHOMSKY NORMAL FORM



- 1. Introduce a new start symbol S_0 , add rule $S_0 \rightarrow S$ (S=old start symbol)
- 2. Eliminate all ε rules of the form $A \rightarrow \varepsilon$ ($A \neq S_0$): remove rule and split rules containing A on the RHS in all versions, with and without A's. For rules $B \rightarrow A$, replace A with ε if B has not been through this step yet, otherwise eliminate $B \rightarrow A$.
- 3. Eliminate all unit rules $A \rightarrow B$, by adding all $B \rightarrow R_i$ to $A \rightarrow R_i$ where R_i is not a unit rule. If R_i is a unit rule add all $R_i \rightarrow K_i$ to A $(A \rightarrow K_i)$ where K_i is not a unit rule. Continue this process for all following unit-rules, until we observe a unit rule we have seen in the cleaning step. Then eliminate $A \rightarrow B$.
- 4. Clean up remaining rules: For $A \rightarrow R_1$, R_2 , ... R_n (n>2, R_i terminals or nonterminals), create a chain $\{A \rightarrow R_1 A_1, A_1 \rightarrow R_2 A_2 ... A_{n-2} \rightarrow R_{n-1} R_n\}$. For all R_i that are terminals, create a lexicon rule and replace R_i with its LHS.
- 5. If $S_0 \rightarrow C$ remains, set C as start symbol.

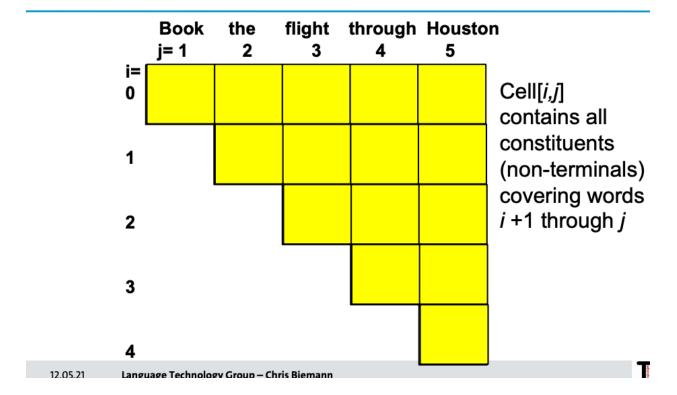
b) Parse the sentence "I see the man with the telescope using the bottom up **CYK" algorithm (convert to CNF)

Assume the following grammar:

 $S \rightarrow NPVP \mid NP$ $VP \rightarrow VNP \mid VNPPP$ $NP \rightarrow NPPP \mid DN \mid R$ $PP \rightarrow PNP$ $P \rightarrow with \mid in \mid of$ $V \rightarrow see \mid take \mid make$ $N \rightarrow man \mid telescope$ $D \rightarrow the \mid an \mid a$ $R \rightarrow I$

CYK PARSER





c) Given the following grammar and lexicon rules, parse the sentence "fruit flies like a banana" with probabilistic CYK.

Grammar

$$S \rightarrow NP \ VP \ 1.0$$

 $NP \rightarrow D \ N \ 0.5$
 $NP \rightarrow NP \ PP \ 0.2$
 $VP \rightarrow VP \ 0.6$
 $VP \rightarrow VP \ 0.4$
 $PP \rightarrow PNP \ 1.0$

Lexicon

$$N \rightarrow fruit 0.2$$
 $V \rightarrow flies 0.1$
 $N \rightarrow flies 0.1$
 $V \rightarrow like 0.3$
 $P \rightarrow like 0.2$
 $D \rightarrow a 0.4$
 $D \rightarrow the 0.5$
 $N \rightarrow banana 0.1$
 $N \rightarrow time 0.1$

Hints from lecture Slides:

PROBABILISTIC CONVERSION TO CNF



Original Grammar		Chomsky Normal Form	
$S \rightarrow NP VP$	0.8	$S \rightarrow NP VP$	0.8
$S \rightarrow Aux NP VP$	0.1	$S \rightarrow X1 VP$	0.1
		$X1 \rightarrow Aux NP$	1.0
$S \rightarrow VP$	0.1	S → book include prefer	
		0.01 0.004 0.006	
		$S \rightarrow Verb NP$	0.05
		$S \rightarrow VP PP$	0.03
$NP \rightarrow Pronoun$	0.2	$NP \rightarrow I \mid he \mid she \mid me$	
		0.1 0.02 0.02 0.06	
NP → Proper-Noun	0.2	NP → Houston NWA	
		0.16 .04	
NP → Det Nominal	0.6	NP → Det Nominal	0.6
Nominal → Noun	0.3	Nominal → book flight meal money	
		0.03 0.15 0.06 0.06	
Nominal → Nominal Noun	0.2	Nominal → Nominal Noun	0.2
Nominal → Nominal PP	0.5	Nominal → Nominal PP	0.5
$VP \rightarrow Verb$	0.2	VP → book include prefer	
		0.1 0.04 0.06	
$VP \rightarrow Verb NP$	0.5	$VP \rightarrow Verb NP$	0.5
$VP \rightarrow VP PP$	0.3	$VP \rightarrow VP PP$	0.3
PP → Prep NP	1.0	PP → Prep NP	1.0

Hints from lecture Slides:

First Steps

PROBABILISTIC CYK PARSING



	Book	the	flight	through	Houston
	S:.01, VP:.1, Verb:.5 Nominal:.03 Noun:.1	None			
S → NP VP S → X1 VP X1 → Aux NP S → book include prefe	0.8 0.1 1.0	Det:.6	NP:.6*.6*.15 =.054		
0.01 0.004 0.00 S → Verb NP S → VP PP NP → I he she me 0.1 0.02 0.02 0.0 NP → Houston NWA	6 0.05 0.03	x → does 1.0	Nominal:.15 Noun:.5		
0.16 .04 NP → Det Nominal Nominal → book flight n 0.03 0.15 (Nominal → Nominal Noun	0.6 neal money _{Pro} 0.06 0.06	t → the a tha 0.6 0.2 0.1 noun → I he 0.5 0.1 0	0.1 she me		
Nominal → Nominal PP VP → book include pref	0.5 Ver er 06 No 0.5	rb → book include 0.5 0.2 un → book flight 0.1 0.5 oper-Noun → Hou	0.3 t meal money 0.2 0.2	,	

Next Step

PROBABILISTIC CYK PARSING



	Book	the	flight	through	Houston
	S:.01, VP:.1, Verb:.5 Nominal:.03 Noun:.1	None	S:.05*.5*.054 =.00135 VP:.5*.5*.054 =.0135		
$S \rightarrow X1 VP$	0.8 0.1 1.0	Det:.6	NP:.6*.6*.15 =.054		
0.01 0.004 0.006 S → Verb NP S → VP PP NP → I he she me 0.1 0.02 0.02 0.0	0.05 0.03	→ does	Nominal:.15 Noun:.5		
Nominal → book flight m 0.03 0.15 0 Nominal → Nominal Noun Nominal → Nominal PP VP → book include prefe 0.1 0.04 0.0 VP → Verb NP	0.6 money Proi 0.06 0.06 0.06 0.0 0.5 Veri 0.5 0.5 0.5 0.5	→ the a tha 0.6 0.2 0.1 noun → I he 0.5 0.1 0 b → book inclu 0.5 0.2 in → book fligh	0.1 she me 0.1 0.3 de prefer 0.3 t meal money 0.2 0.2	,	

PROBABILISTIC CYK PARSING



Book	<	the	flight	through	Houstor
S:.01, VI Verb:.5 Nominal Noun:.1	- 1	None	S:.05*.5*.054 =.00135 VP:.5*.5*.054 =.0135	None	
S → NP VP 0.8 S → X1 VP 0.1 X1 → Aux NP 1.0 S → book include prefer		Det:.6	NP:.6*.6*.15 =.054	None	
0.01 0.004 0.006 S → Verb NP 0.05 S → VP PP 0.03 NP → I he she me 0.1 0.02 0.02 0.06 NP → Houston NWA	Aux	→ does	Nominal:.15 Noun:.5	None	
0.16 .04 NP → Det Nominal 0.6 Nominal → book flight meal moner 0.03 0.15 0.06 0.06 Nominal → Nominal Noun 0.2	y Pror	→ the a tha 0.6 0.2 0.1 noun → I he 0.5 0.1 0 → book inclu	0.1 she me .1 0.3	Prep:.2	
$\begin{array}{lll} \mbox{Nominal} \rightarrow \mbox{Nominal} \mbox{PP} & 0.5 \\ \mbox{VP} \rightarrow \mbox{book} \mid \mbox{include} \mid \mbox{prefer} \\ 0.1 & 0.04 & 0.06 \\ \mbox{VP} \rightarrow \mbox{Verb} \mbox{NP} & 0.5 \\ \mbox{VP} \rightarrow \mbox{VP} \mbox{PP} & 0.3 \\ \end{array}$	Nou	0.5 0.2	0.3 t meal money 0.2 0.2		

Final Step

SIMPLE PCFG FOR A SUBSET OF ENGLISH



Grammar P	rob. Lex	icon
S → NP VP S → Aux NP VP S → VP NP → Pronoun NP → Proper-Noun NP → Det Nominal Nominal → Noun	0.8 0.1 0.1 0.2 0.2 0.2 0.6 0.3	Det → the a that this
Nominal → Nominal Noun Nominal → Nominal PP VP → Verb VP → Verb NP VP → VP PP PP → Prep NP	0.3 0.2 0.5 0.2 0.5 1.0 1.0	$0.5 \ 0.1 \ 0.1 \ 0.3$ Proper-Noun \rightarrow Houston NWA $0.8 \ 0.2$ Aux \rightarrow does 1.0 Prep \rightarrow from to on near through $0.25 \ 0.25 \ 0.1 \ 0.2 \ 0.2$

PROBABILISTIC CYK PARSING

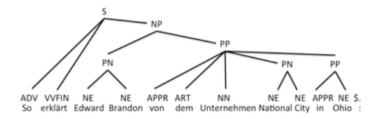


	Book	the	flight	through	Houstor	1
	S:.01, VP:.1, Verb:.5 Nominal:.03 Noun:.1	None	S:.05*.5*.054 =.00135 VP:.5*.5*.054 =.0135	None	-S:.0000216	Pick most probable parse, i.e. take max to combine probabilities
	0.8 0.1 1.0	Det:.6	NP:.6*.6*.15 =.054	None	NP:.6*.6* .0024 =.000864	of multiple derivations of each constituent in each cell.
0.01 0.004 0.000 S → Verb NP	0.05 0.03	a → does	Nominal:.15 Noun:.5	None	Nominal: .5*.15*.032 =.0024	
0.16 .04 NP → Det Nominal Nominal → book flight m 0.03 0.15 0 Nominal → Nominal Noun	0.6 neal money Proj 0.06 0.06	→ the a tha 0.6 0.2 0.1 noun → I he 0.5 0.1 0 b → book inclu	0.1 she me .1 0.3	Prep:.2 ←	PP:1.0*.2*.16 =.032	
VP → book include prefe 0.1 0.04 0.0 VP → Verb NP	er 06 Nou 0.5	0.5 0.2 un → book fligh 0.1 0.5 per-Noun → Hou		,	NP:.16 PropNoun:.8	

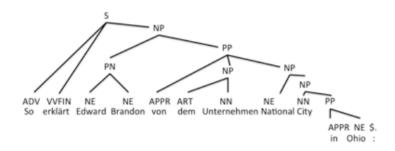
Type $\it Markdown$ and LaTeX: $\it \alpha^2$

a) Compute Precision, Recall and F1 value for the following two parses, labeled and unlabeled.

System:



Gold:



b) Convert the trees in a) into the Penn Treebank bracketed format.

Hints from Lecture Slides:

```
💾 Universität Hamburg
FORMAT
Every production rule is ((S
represented by
                             (NP-SBJ (DT The) (NNP Illinois) (NNP Supreme) (NNP Court))
                             (VP (VBD ordered)
• (
                              (NP-1 (DT the) (NN commission))
  left hand side
                               (NP-SBJ (-NONE- *-1) )
   sequence of right
                               (VP (TO to)
   hand side symbols
                                (VP
                                 (VP (VB audit)
      - non-terminals
       expanded by
                                  (NP
                                   (NP (NNP Commonwealth) (NNP Edison) (POS 's) )
        production rule
                                   (NN construction) (NNS expenses) ))
       terminals
                                 (CC and)
                                 (VP (VB refund)
                                  (NP (DT any) (JJ unreasonable) (NNS expenses) ))))))
                             (..)))
Traces: -NONE- and
trace-number
```

c) Familiarize with the Stanford CoreNLP Parser https://corenlp.run/) or http://nlp.stanford.edu:8080/parser/). Parse a sentence and draw the tree (using constituency parse). Does your sentence make sense?