Onderwerp:

Datum:

Week 2

2) Trigram probability $= P(w_n | w_{n-2} w_{n-1}) = \frac{C(w_n w_{n-1} w_{n-2})}{C(w_{n-2} w_{n-1})}$

$P(\text{Sam} | \text{I am}) = 0.5$

3) P(<s> i want chinese food </s>) =

$P(i | \langle s \rangle) P(\text{want} | i) P(\text{chinese} | \text{want}) P(\text{food} | \text{chinese}) P(\langle /s \rangle | \text{food})$

$= 0.25 \times 0.33 \times 0.0065 \times 0.52 \times 0.68 = 1{,}896 \times 10^{-4}$

↳ according to Figure 3.2

$= 0.19 \times 0.21 \times 0.0029 \times 0.052 \times 0.040 = 2{,}406 \times 10^{-6}$

↳ according to Figure 3.7

4) From Equation 3.17 from the book:

$$PP(W) = P(w_1 w_2 \ldots w_N)^{-1/N}$$

For training set, unigram perplexity is:

$$PP(w) = \left[\left(\frac{91}{100} \right)^{91} \left(\frac{1}{100} \right)^{9} \right]^{-1/100} = 1{,}64919$$

For test set, unigram perplexity is:

$$PP(w) = \left[\left(\frac{9}{10} \right)^{9} \left(\frac{1}{10} \right)^{1} \right]^{-1/10} = 1{,}38415$$

* It was expected that perplexity of the test set is going to be lower compared to calculation followed by Equation 3.17 since mostly, the next number will be zero.

Onderwerp:

Datum:

Week 3

1)

1. [a-zA-Z]+
2. [a-z]*b
3. (b+(ab+)+)?

2)

1. ([a-zA-Z]+)\s+\1
2. ^\d+.*\b[a-zA-Z]+$
3. \bgrotto\b.*\braven\b|\braven\b.*\bgrotto\b

Week 4

2)

1. promise/NOUN → it should be VERB.
2. that/DET → it should be SCONJ.
3. twenty/DET → it should be NUM.
4. quickly/ADJ → it should be ADV.
5. John/NOUN → it should be PROPN.

gratis collegeblok

Onderwerp:

Datum:

Week 4:

3)

Transition probability table

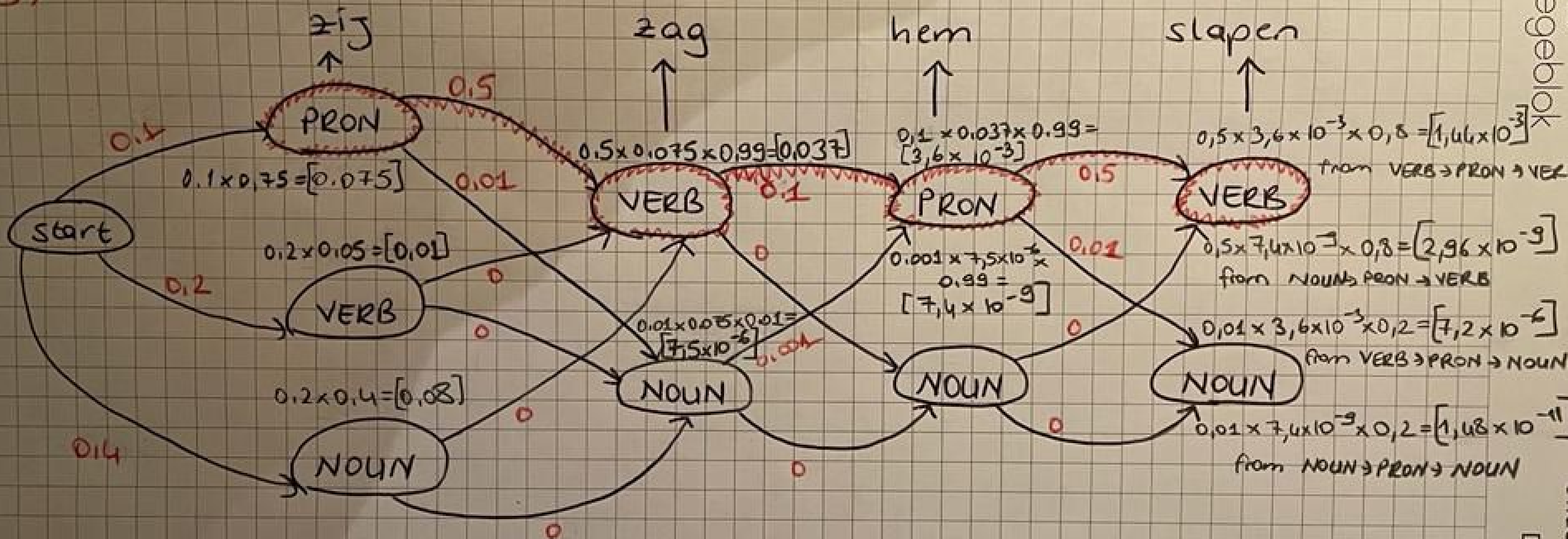
	DET	NOUN	PRON	VERB	END
start	0.1	0.4	0.1	0.2	—
DET	—	0.4	—	0.00001	0.01
PRON	—	0.01	—	0.5	0.33
VERB	0.01	—	0.1	—	0.33
NOUN	0.001	—	0.001	—	0.33

Emission probability table

	zij	zag	hem	slapen
PRON	0.75	—	0.99	—
VERB	0.05	0.99	—	0.8
NOUN	0.2	0.01	0.01	0.2

Week 4

3)



By Viterbi Algorithm:

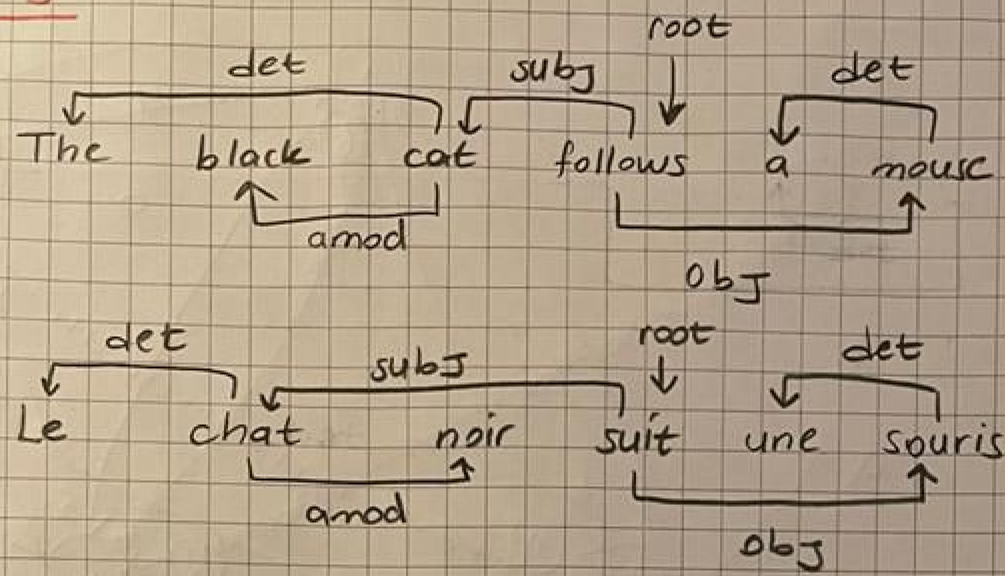
zij	zag	hem	slapen
PRON	VERB	PRON	VERB
$[7{,}5\times10^{-2}]$	$[3{,}7\times10^{-2}]$	$[3{,}6\times10^{-3}]$	$[1{,}44\times10^{-3}]$

Onderwerp:

Datum:

Week 5

1)



2) When we assume above is golden:

$$UAS = \frac{9-2}{9} = \frac{7}{9} \qquad LAS = \frac{7-2}{9} = \frac{5}{9}$$

When we assume below is golden:

$$UAS = \frac{8-3}{8} = \frac{5}{8} \qquad LAS = \frac{5-2}{8} = \frac{3}{8}$$

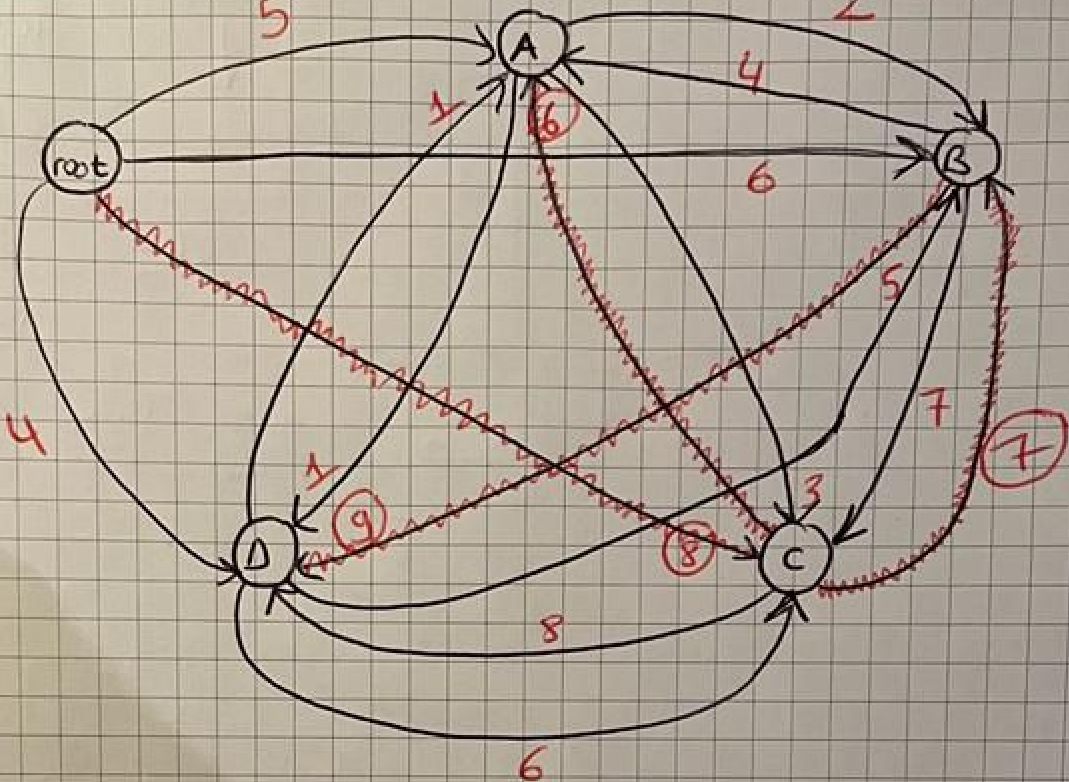
Since $UAS = \frac{\text{nodes with correct parent}}{\text{total nodes}}$ and

$LAS = \frac{\text{nodes with correct parent and edge label}}{\text{total nodes}}$

Datum:

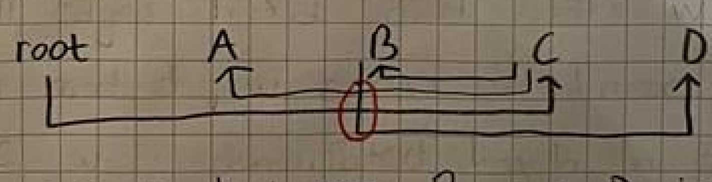
Week 5

3)



1. Yes, it is a valid dependency tree (see chapter 14.2 from the book - Dependency Formalisms. All the three constraints that are presented hold for the above graph).
   (first step of the CLE algorithm is edge values that are marked with "O").

2. The graph drawn above is not projective.

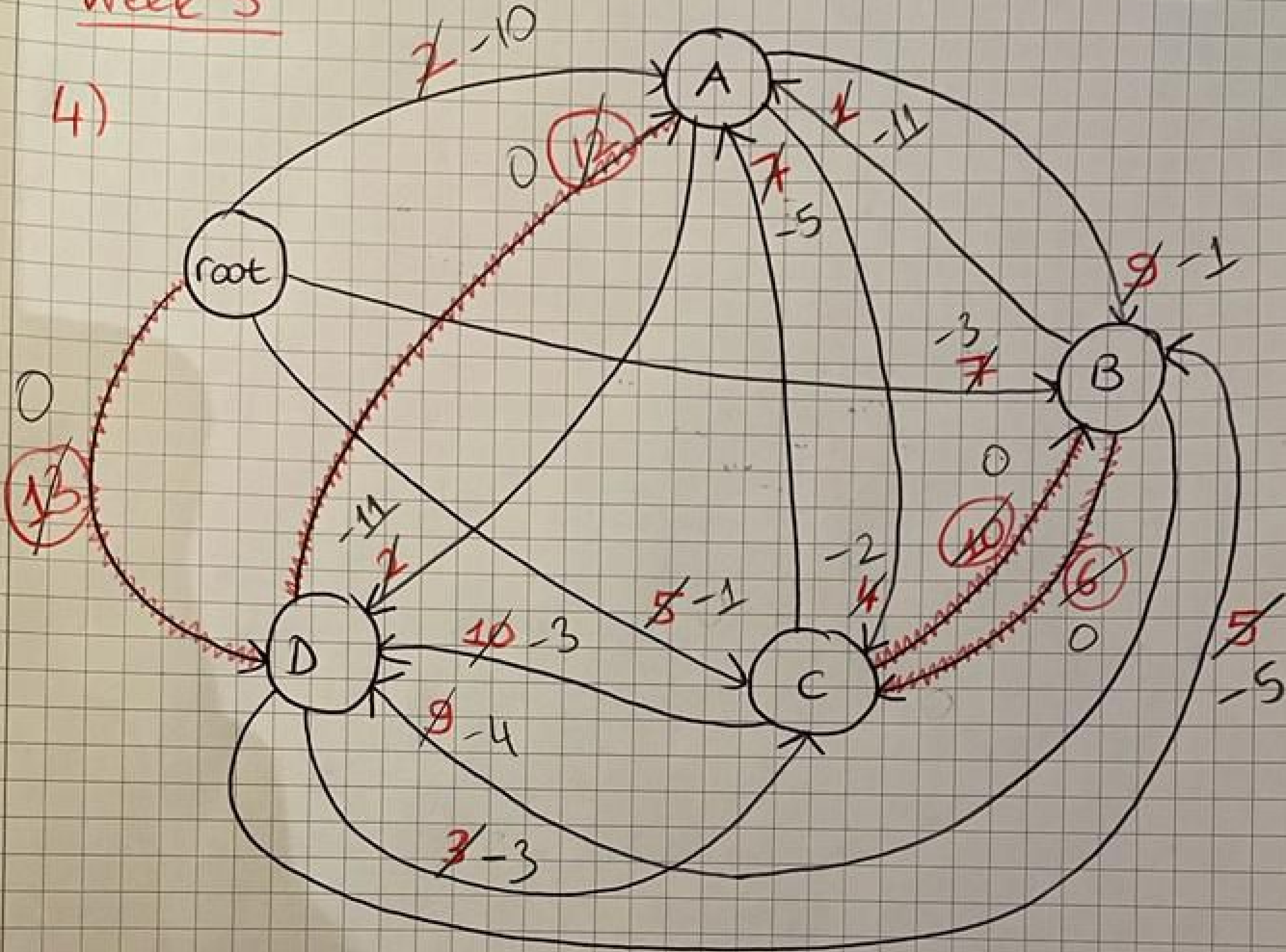


C occurs between B and D in linear order without being dominated by B. (see week 5 slides, page 16 - projectivity).

3. There is no need to continue with CLE algorithm since above graph is already a valid dependency tree.

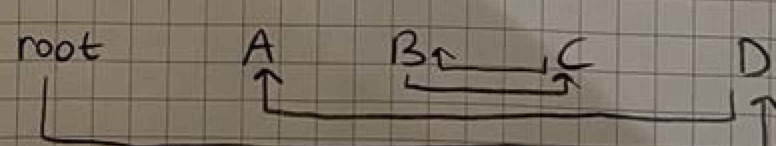
Week 5

4)



1. Above graph is not a valid dependency tree since there is a cycle between B and C.

2. It is not projective since there is no path from the head to every word that lies between the head and the dependent. (for example D → A, no path from B and C to A).



3. Next step in the CLE algorithm is subtracting the best incoming edge score from all incoming edges for a node. Calculations can be seen in above graph.