ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE
EIDGENÖSSISCHE TECHNISCHE HOCHSCHULE – LAUSANNE
POLITECNICO FEDERALE – LOSANNA
SWISS FEDERAL INSTITUTE OF TECHNOLOGY – LAUSANNE

Faculté Informatique et Communication
Introduction to Natural Language Processing (Ms; CS-431)
Chappelier, J.-C. & Rajman, M.

INTRODUCTION TO NATURAL LANGUAGE PROCESSING

Solution and scoring scale to Fall 2020, Graded Quiz #3

NAME: **SCIPER:**

Instructions:

You have 45 minutes (8:15–9:00) for this quiz, which consists of several questions of different weights. For each question the corresponding points are indicated and the total number of points is 30.
[...]

The solution comes in red. And some comments in blue. And grading scale in green.

QUESTION I [3 pt]

Consider the following lexicon L:

```
a        : Det
blue     : Adj, N
drink    : N, V
drinks   : N, V
friends  : N
from     : Prep
gave     : V
letter   : N
my       : Det
neighbor : N
nice     : Adj, N
of       : Prep
postman  : N
ran      : V
the      : Det
to       : Prep
```

① **[1 pt]** When using an order-1 HMM model to tag the word sequence
"*nice friends drink blue drinks*",
does the tagging of "*drink*" depend on the one of "*nice*"?

(Select only one answer.)

[ ] yes, because the HMM approach relies on a global maximum
[✔] no, the used hypotheses make the two taggings independent from each other
since "*friends*" is not ambiguous and we use a order-**1** HMM.

② **[2 pts]** When using an order-2 HMM model based on the lexicon L, what is the number of *distinct* parameters that are used for tagging the word sequence
"*the nice postman gave the blue letter*"?

We assume that a smoothing mechanism is used that guarantees that none of these parameters can be zero.

(Select only one answer.)

[ ] it depends on how the parameters are estimated
[ ] 13
[✔] 17
2 initial: $P(\texttt{Det Adj})$, $P(\texttt{Det N})$;
8 emission: $P(\text{the}|\texttt{Det})$, $P(\text{nice}|\texttt{Adj})$, $P(\text{nice}|\texttt{N})$, $P(\text{postman}|\texttt{N})$, $P(\text{gave}|\texttt{V})$, $P(\text{blue}|\texttt{Adj})$, $P(\text{blue}|\texttt{N})$, $P(\text{letter}|\texttt{N})$;
and 7 transition: $P(\texttt{N}|\texttt{Det Adj})$, $P(\texttt{N}|\texttt{Det N})$, $P(\texttt{V}|\texttt{Adj N})$, $P(\texttt{V}|\texttt{N N})$, $P(\texttt{Det}|\texttt{N V})$, $P(\texttt{N}|\texttt{V Det})$, $P(\texttt{Adj}|\texttt{V Det})$.
[ ] 20
Incorrect because it is counting *twice* the parameters $P(\text{the}|\texttt{Det})$, $P(\texttt{N}|\texttt{Det Adj})$ and $P(\texttt{N}|\texttt{Det N})$.

QUESTION II [2 pt]

Which of the following are *parameters* involed in the choice made by an order-1 HMM model for PoS tagging knowing that its output is

```
this/Pron is/V a/Det good/Adj question/N
```

and that neither 'is' nor 'question' can be adjectives, and 'question' cannot be a determiner.

(Select one or several answers. Penalty for wrong ticks.)

[-100%] P(N|question)
[✔33%] P(question|N)
[-100%] P(question|Adj N)

[-100%] P(question|N Adj)

[-100%] P(this)

[-100%] P(this is)

[-100%] P(this V)

[✔33%] P(Pron)

[-100%] P(Pron V)

[-100%] P(Pron is)

[-100%] P(Det|Adj)

[✔33%] P(Adj|Det)

[-25%] P(Adj|V Det)

[-100%] P(Adj|Det V)

[-100%] P(Det|V Adj)

[-100%] P(Det|Pron V)

[-100%] P(Adj|a)

[-100%] P(question|Adj)

QUESTION III [7 pt]

We consider the input sequence

process time control

and an order-1 Markov Model for PoS tagging with the following parameters, all expressed here up to some (unique same) multiplicative constant:

initial:

N	V	Adj
3	2	1

emission (where an hyphen (–) means 0, "impossible"):

	process	time	control
N	4	7	8
V	7	10	9
Adj	—	5	—

transition $P(Y|X)$:

		Y		
		N	V	Adj
X	N	12	11	7
	V	5	6	3
	Adj	8	4	2

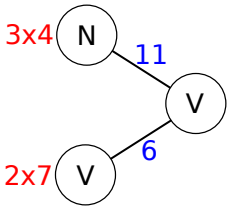
① **[1 pt]** For each column of the lattice used in the Viterbi algorithm applied to that input sequence ("*process time control*"), how many nodes (= lines) does it have?

Provide your answer as a coma separated list of integers. For instance, if the lattice has 5 columns, you could for instance here write: 4,3,2,4,2

Answer: 2, 3, 2

② **[3 pts]** What value would be stored by the Viterbi algorithm in the node associated to V for the word "*time*"?

Answer: 1320
$= P(N) \cdot P(\text{process}|N) \cdot P(N|N) \cdot P(\text{times}|V)$
$= 3 \times 4 \times 11 \times 10$
since $3 \times 4 \times 11 > 2 \times 7 \times 6$ (V for "*process*").

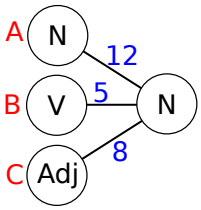


③ **[3 pts]** Let's denote by respectively A, B and C the value stored by the Viterbi algorithm in the node associated to respectively N, V and Adj for the word "*time*".

If $C > B > A$ and $10A \geq 9C$, what would be the tag of "*time*" in the most probable tagging, if the tag of "*control*" is N (in the most probable tagging)?

Answer: N
since:

$$12A > 10A \geq 9C > 8C > 8B > 5B$$



Note: B is indeed the value computed before: 1320.

QUESTION IV [13 pt]

Consider the following context-free grammar G (where `S` is the top-level symbol):

```
R01: S   --> NP VP
R02: NP  --> NP0
R03: NP  --> Det NP0
R04: NP0 --> N
R05: NP0 --> Adj N
R06: NP0 --> NP0 PNP
R07: VP  --> V
R08: VP  --> V NP
R09: VP  --> V NP PNP
```

```
R10: PNP --> Prep NP
```

complemented by the lexicon L previously defined, repeated here for convenience:

```
a        : Det
blue     : Adj, N
drink    : N, V
drinks   : N, V
friends  : N
from     : Prep
gave     : V
letter   : N
my       : Det
neighbor : N
nice     : Adj, N
of       : Prep
postman  : N
ran      : V
the      : Det
to       : Prep
```

① **[1 pt]** Indicate which of the following statements are true for G?

(Select one or several answers. Penalty for wrong ticks.)

[✔50%] G recognizes an infinite number of (syntactically) acceptable word sequences.

[-100%] G is simple enough for any word sequence to be parsed in a linear time (wrt its length).

[✔50%] G is equivalent to a grammar in Chomsky Normal Form.
(as this is the case for *any* CFG)

[-50%] G cannot be used as such with the bottom-up chart parsing algorithm.
(it is precisely the purpose of the bottom-up chart parsing algorithm to deal with *any* CFG)

② **[1 pt]** Indicate the number of non-terminals contained in the grammar G: 10: 5 pre-terminals and 5 other non-terminals.

③ **[2.5 pt]** Which of the following word sequences are syntactically correct according to G?

(Select one or several answers. Penalty for wrong ticks.)

[✔60%] the letter of the neighbor to a blue from the friends gave

[✔40%] a nice ran postman of my blue

[-100%] a neighbor postman drinks
(there is no way to have two nouns one after the other)

④ **[2.5 pt]** If the notation "$T(w)$" is used to refer to the rule "$T \longrightarrow w$", which of the following correspond to valid derivations according to the grammar G?

(Select one or several answers. Penalty for wrong ticks.)

[-100%] [R01, R08, R02, R04, N(letter), V(ran), R03, Det(the), R04, N(drinks)]
(it does not follow the left-most-non-terminal-first rewrite convention)
[✔100%] [R01, R03, Det(a), R05, Adj(blue), N(drink), R07, V(ran)]
[-100%] [R01, R02, R04, N(friends), R09, V(gave), R02, N(postman)]
(R04 is missing before N(postman))

⑤ **[3 pts]** How many parse trees does the grammar G associate to the word sequence "*the postman ran the letter for the drinks on the friends*"?

(Select only one answer.)

[ ] 0
[ ] 1
[ ] 2
[✔] 4
[ ] 8

There is no point doing here, by hand, a CYK chart: (we are not computers:) the sentence is too long and the structure of the grammar should be simple enough to understand it:

- since "*drinks*" is preceded by "*the*", only "*ran*" can be `V`;
- thus, to be a correct sentence, "*the postman*" must be `NP` and "*ran the letter for the drinks on the friends*" shall be `VP`;
- there are only 2 (non-unary) rules the left-hand side of which is `VP` to be considered:
  - `R08`, which has only 2 possibilities (`PNP` attachment):
    - "*drinks*" and "*friends*" attached to "*letter*"
    - "*friends*" attached to "*drinks*" and "*drinks*" attached to "*letter*"
  - `R09`, which also has only 2 possibilities (`PNP` attachment):
    - "*drinks*" attached to "*letter*" only (since we need another `PNP` to be used in `R09`)
    - "*friends*" attached to "*drinks*" only

⑥ **[1 pt]** Indicate which of the following statements are true for the word sequence W = "*drinks drinks drinks*"?

(Select only one answer.)

[ ] W is not syntactically acceptable according to the grammar G.
[ ] W is syntactically acceptable according to the grammar G, but some *positional* constraints of English are violated.
[✔] W is syntactically acceptable according to the grammar G, but some *selectional* constraints of English are violated.
[ ] W is syntactically acceptable according to the grammar G, but some *positional* **and** *selectional* constraints of English are violated.

⑦ **[2 pt]** How many (syntactic and lexical) rules does the extended Chomsky Normal Form grammar equivalent to G contain, if produced as described in the parsing lecture?

(Select only one answer.)

[ ] the grammar G cannot be converted to extended Chomsky Normal Form
[ ] the grammar G already is in extended Chomsky Normal Form
[50%] 11 rules
(don't forget lexical rules)
[✔] 31 rules
[ ] 48 rules

There are 30 rules in the original grammar (20 lexical and 10 non-lexical), only one of which must be transformed (R09), which is done by replacing it with two new rules.

QUESTION V **[5 pt]**

Consider the following CFG

```
S    --> NP VP PNP
NP   --> Det N
NP   --> Det Adj N
VP   --> V
VP   --> Aux Ving
VP   --> VP NP
VP   --> VP PNP
PNP  --> Prep NP
```

and the following lexicon:

```
the:Det, red:Adj, cat:N, is:Aux, meowing:Ving, on:Prep, roof:N
```

The next four questions ask you the content of a given cell of the chart used by the CYK algorithm (used here as a recognizer) for the input sentence

the red cat is meowing on the roof

Prior to answering the question, we must transform the grammar to CNF. Two rules are concerned:

```
S    --> NP VP PNP
NP   --> Det Adj N
```

each of which could be transformed two ways:

```
X1   --> NP VP
S    --> X1 PNP
```

or

```
Z1   --> VP PNP
S    --> NP Z1
```

and

```
X2   --> Det Adj
NP   --> X2 N
```

or

```
Z2   --> Adj N
NP   --> Det Z2
```

① **[1 pt]** What is the content of the cell at row 3 column 1 (indexed as in the lectures)?

NP (corresponds to "*the red cat*")

② **[1 pts]** What is the content of the cell at row 3 column 6 (indexed as in the lectures)?

PNP (corresponds to "*on the roof*")

③ **[2 pts]** Knowing that cell at row 5 column 4 (indexed as in the lectures) contains **only** a VP, what is the content of the cell at row 5 column 1 ?

X1 (corresponds to "*the red cat is meowing*"): since cell $(5, 4)$ contains only a VP and does not contain Z1, the first rule was transformed using X1

④ **[1 pts]** What is the content of the cell at row 8 column 1 (indexed as in the lectures)?

S (50%) and X1 (since the PNP could be inside the VP)

Notice that S is present only **ONCE**, even though there are 2 (full) parse trees.

For those who prefer to see the full CYK chart, here it is:

8	S, X1							
7								
6								
5	X1			VP				
4								
3	NP					PNP		
2	X2			VP			NP	
1	Det	Adj	N	Aux	Ving	Prep	Det	N
	1	2	3	4	5	6	7	8
	the	red	cat	is	meowing	on	the	roof