

TP – classification en POS

- 3 modèles
 - « mots »
 - mots + contextes gauche et droit
 - « mots+morpho »
 - mots + contextes gauche et droit
 - n-grams de lettres pour le mot et ses contextes droits et gauches
 - « mots+morpho+lexique »
 - mots + contextes gauche et droit
 - n-grams de lettres pour le mot et ses contextes droits et gauches
 - liste de POS possible selon un lexique (Lefff)

Exemple

```
1 , XX , XX , XX , XX , certes , # c e r t e s # , !VIRGULE , !VIRGULE , rien , # r i e n # , adv , adv .
2 , XX , XX , certes , # c e r t e s # , !VIRGULE , !VIRGULE , rien , # r i e n # , ne , # n e # , ponctw , ponctw .
3 , certes , # c e r t e s # , !VIRGULE , !VIRGULE , rien , # r i e n # , ne , # n e # , dit , # d i t # , nc pro , pro .
4 , !VIRGULE , !VIRGULE , rien , # r i e n # , ne , # n e # , dit , # d i t # , qu' , # q u ' # , OOV , advneg .
5 , rien , # r i e n # , ne , # n e # , dit , # d i t # , qu' , # q u ' # , une , # u n e # , adj nc v , v .
6 , ne , # n e # , dit , # d i t # , qu' , # q u ' # , une , # u n e # , seconde , # s e c o n d e # , prrel pri , csu .
7 , dit , # d i t # , qu' , # q u ' # , une , # u n e # , seconde , # s e c o n d e # , motion , # m o t i o n # , det nc pro , det .
8 , qu' , # q u ' # , une , # u n e # , seconde , # s e c o n d e # , motion , # m o t i o n # , de , # d e # , adj adj nc nc v , adj .
9 , une , # u n e # , seconde , # s e c o n d e # , motion , # m o t i o n # , de , # d e # , censure , # c e n s u r e # , nc , nc .
10 , seconde , # s e c o n d e # , motion , # m o t i o n # , de , # d e # , censure , # c e n s u r e # , sur , # s u r # , det prep , prep .
11 , motion , # m o t i o n # , de , # d e # , censure , # c e n s u r e # , sur , # s u r # , son , # s o n # , nc v , nc .
12 , de , # d e # , censure , # c e n s u r e # , sur , # s u r # , son , # s o n # , projet , # p r o j e t # , adj prep , prep .
13 , censure , # c e n s u r e # , sur , # s u r # , son , # s o n # , projet , # p r o j e t # , de , # d e # , det nc , det .
14 , sur , # s u r # , son , # s o n # , projet , # p r o j e t # , de , # d e # , loi , # l o i # , nc , nc .
15 , son , # s o n # , projet , # p r o j e t # , de , # d e # , loi , # l o i # , !VIRGULE , !VIRGULE , det prep , prep .
16 , projet , # p r o j e t # , de , # d e # , loi , # l o i # , !VIRGULE , !VIRGULE , reprenant , # r e p r e n a n t # , nc , nc .
17 , de , # d e # , loi , # l o i # , !VIRGULE , !VIRGULE , reprenant , # r e p r e n a n t # , l' , # l ' # , ponctw , ponctw .
18 , loi , # l o i # , !VIRGULE , !VIRGULE , reprenant , # r e p r e n a n t # , l' , # l ' # , accord , # a c c o r d # , v , v .
```

id:	11 ,
moins2mot:	motion ,
moins2let:	# m o t i o n # ,
moins1mot:	de ,
moins1let:	# d e # ,
mot:	censure ,
let:	# c e n s u r e # ,
plus1mot:	sur ,
plus1let:	# s u r # ,
plus2mot:	son ,
plus2let:	# s o n # ,
cate_poss_lefff:	nc v ,
label:	nc .

Modele 1 : traits lexicaux

```
id: ignore.  
moins2mot: ignore.  
moins2let: ignore.  
moins1mot: text.  
moins1let: ignore.  
mot: text.  
let: ignore.  
plus1mot: text.  
plus1let: ignore.  
plus2mot: ignore.  
plus2let: ignore.  
cate_poss_lefff: ignore.
```

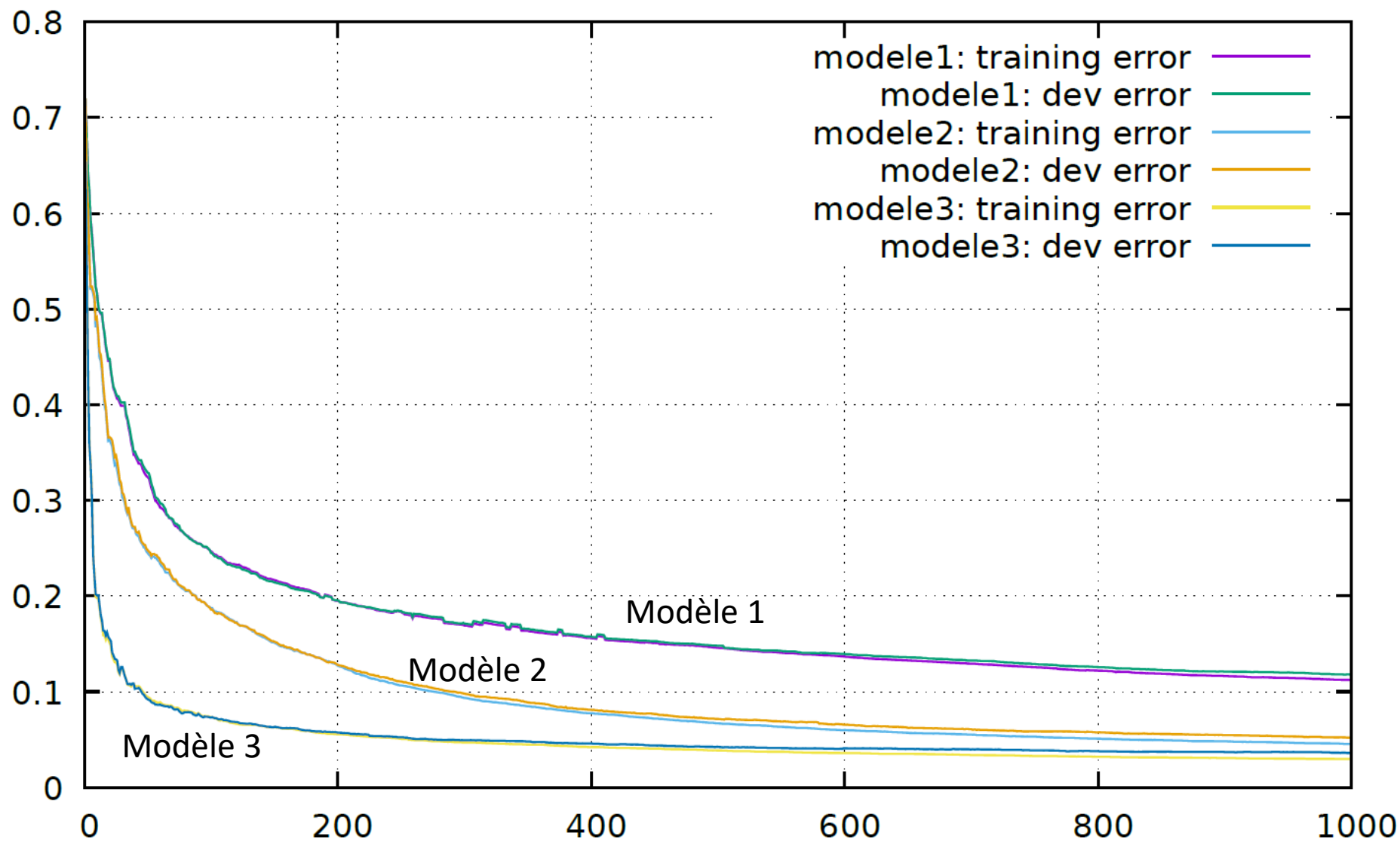
Modele 2 : traits lexicaux + morphologie

```
moins2mot: text.  
moins2let: ignore.  
moins1mot: text.  
moins1let: text: expert_type=ngram expert_length=3  
cutoff=10.  
mot: text.  
let: text: expert_type=ngram expert_length=3 cutoff=10.  
plus1mot: text.  
plus1let: text: expert_type=ngram expert_length=3  
cutoff=10.  
plus2mot: text.  
plus2let: ignore.  
cate_poss_lefff: ignore.
```

Modele 3 : traits lexicaux + morphologie + lexique

```
id: ignore.  
moins2mot: text.  
moins2let: ignore.  
moins1mot: text.  
moins1let: text: expert_type=ngram expert_length=3 cutoff=10.  
mot: text.  
let: text: expert_type=ngram expert_length=3 cutoff=10.  
plus1mot: text.  
plus1let: text: expert_type=ngram expert_length=3 cutoff=10.  
plus2mot: text.  
plus2let: ignore.  
cate_poss_lefff: text: expert_type=ngram expert_length=1 cutoff=2.
```

Courbe d'apprentissage



Sélection de traits par les modèles

Explicite dans les fichiers « XX.shyp »

Modèle 1

```
mot:!VIRGULE
plus1mot:XX
mot:de
mot:et
mot:la
mot:"
mot:l'
mot:le
mot:les
mot:qui
mot:à
mot:_NUM_
plus1mot:!VIRGULE
plus1mot:de
mot:des
mot:en
mot:d'
mot:(
mot:)
moins1mot:la
```

Modèle 2

```
let:#
let:##1
let:t
let:##d
plus1let:#
let:r
let:q#u
moins1let:##1
let:s
let:é
let:##u#n
let:U#M
let:à
let:##e
let:i
mot:!POINT
let:##a
let:e
plus1let:##1
mot:de
```

Modèle 3

```
cate_poss_lefff:det
cate_poss_lefff:ponctw
cate_poss_lefff:prep
cate_poss_lefff:nc
cate_poss_lefff:v
plus1let:XX
cate_poss_lefff:adj
cate_poss_lefff:coo
cate_poss_lefff:adv
cate_poss_lefff:OOV
cate_poss_lefff:clo
cate_poss_lefff:np
cate_poss_lefff:prorel
cate_poss_lefff:cln
let:U#M
cate_poss_lefff:pro
moins1let:##1
let:t
cate_poss_lefff:csu
cate_poss_lefff:nc
```

Différence V1/V2

- Rajout de « bruit » entre V1 et V2
 - des lettres sont permutées aléatoirement pour les verbes, les noms et les adjectifs dans V2
 - Mots inconnus plus important
 - 3797 mots différents dans V1
 - 4712 mots différents dans V2
- Permet de tester les capacités de généralisation des modèles
 - Utilisation du contexte
 - Utilisation de la morphologie
 - Robustesse au « bruit »

1	alors_que	csu
2	m.	nc
3	Louis	np
4	iVannet	np
5	dervait	v
6	usccéder	v
7	à	prep
8	m.	nc
9	Henri	np
10	Karsucki	np
11	au	prep
12	poste	nc
13	de	prep
14	escrétaire	nc
15	géénral	adj
16	,	ponctw
17	la	det
18	moitié	nc
19	du	prep
20	bureau	nc
21	cofnédéral	adj
22	(ponctw
23	seize	det
24	mebmres	nc
25)	ponctw
26	pourarit	v
27	être	v
28	rneouvelée	v
29	.	poncts

Evaluation

- Modele 1

V1 : EXAMPLE ERROR RATE (argmax): 0.069 (1065/15418)

V2 : EXAMPLE ERROR RATE (argmax): 0.109 (1681/15418)

- Modele 2

- V1 : EXAMPLE ERROR RATE (argmax): 0.032 (507/15418)

- V2 : EXAMPLE ERROR RATE (argmax): 0.056 (864/15418)

- Modele 3

- V1 : EXAMPLE ERROR RATE (argmax): 0.029 (458/15418)

- V2 : EXAMPLE ERROR RATE (argmax): 0.069 (1066/15418)

Pourquoi Modele 3 plus mauvais sur V2 ?

Modèle 3

V1

```
EXAMPLE ERROR RATE (argmax): 0.029706 (458/15418)
adj: r=0.933506 p=0.921569 f=0.927499
nc: r=0.971937 p=0.972818 f=0.972377
np: r=0.975986 p=0.931260 f=0.953099
v: r=0.971783 p=0.982877 f=0.977299
```

V2

```
EXAMPLE ERROR RATE (argmax): 0.069140 (1066/15418)
adj: r=0.732297 p=0.819324 f=0.773370
nc: r=0.903440 p=0.923789 f=0.913501
np: r=0.933105 p=0.737127 f=0.823618
v: r=0.903499 p=0.910176 f=0.906825
```