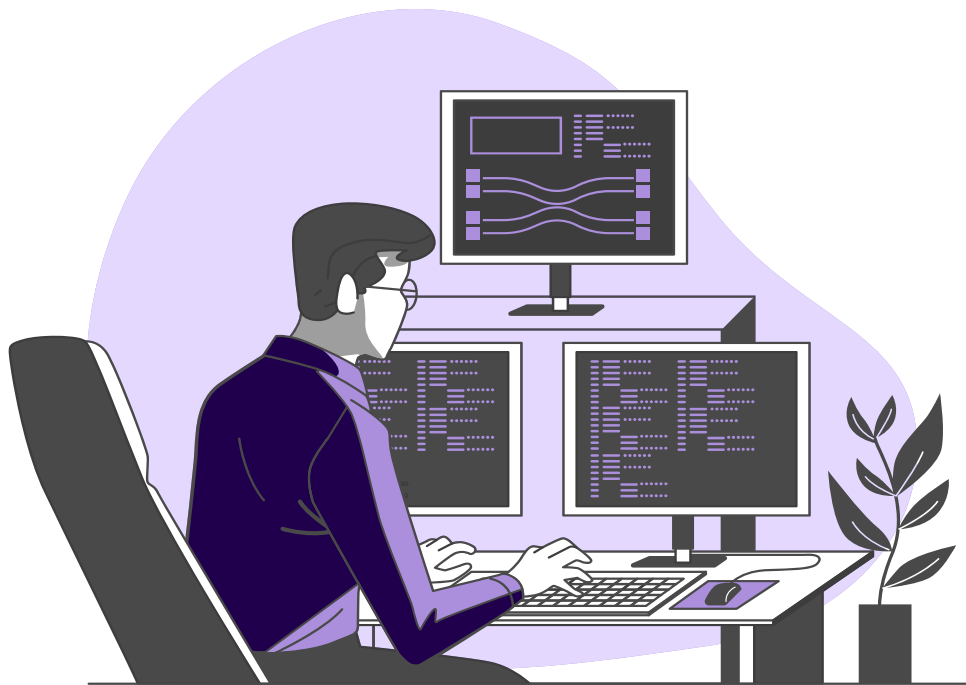


Kenza, Sandra, Shami, Valentina



# Étiquetage en POS avec des ressources faibles

Exemple d'annotation des  
langues peu dotées.



Željko Agić, Dirk Hovy, and Anders Søgaard

**“If all you have is a bit of the Bible:  
Learning POS taggers for truly  
low-resource languages”**



01

## Les langues traités

Une 100-aine de langues (plupart non Indo-Européennes)



02

## Les données

Corpus biblique

03

## La méthodologie

- projection des annotations.
- entraînement semi-supervisé.

04

## Les résultats

- entre 50 et 90% (selon langue)
- moyenne 70% (quand OOV < 50%)

# Langues et Données

- Langues

- Non Indo-Européennes (ex. Akawaio, Aukan, Cakchiquel)
- Peu dotées en ressources.
- 39 ont moins d'un million de locuteurs.

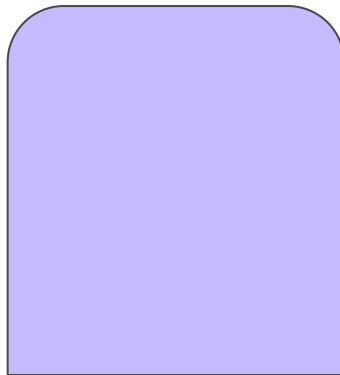
- Données

- Bible comme principale source de données (Corpus Biblique Parallèle Multilingue d'Édimbourg (Christodouloupoulos et Steedman, 2014)).
- 100 traductions.



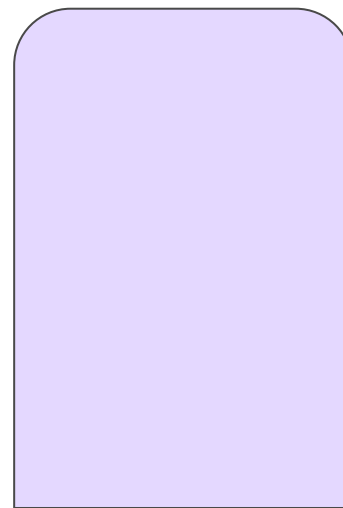
42 langues

Traduction du  
nouveau testament



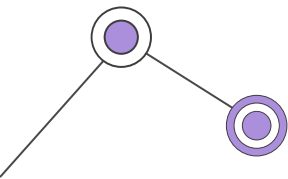
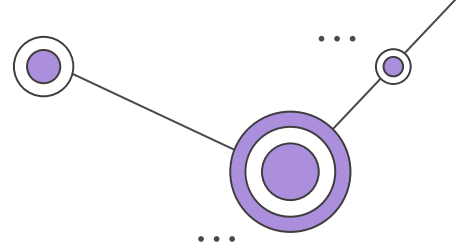
54 langues

Traductions  
complètes de la Bible



4 langues

Extraits de traduction  
du Nouveau Testament



# Méthodologie

01

## Projection d'annotation

- Projection des annotations des Part Of Speech à partir de langues annotées vers non annotées (alignements de tokens entre les traductions..)

02

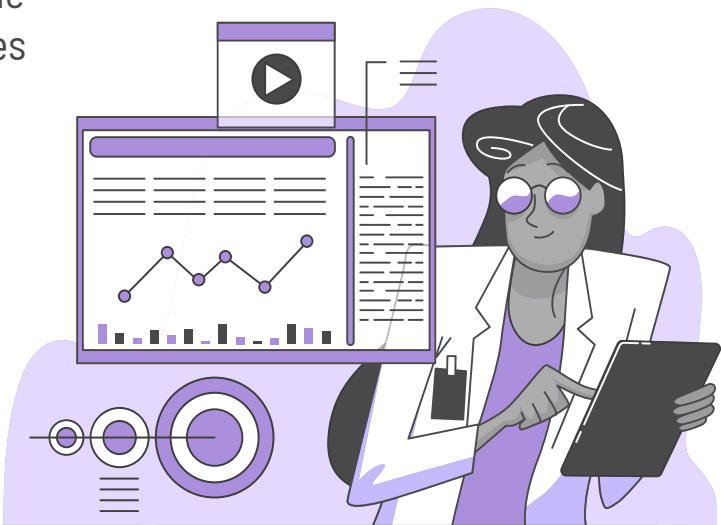
## Entraînement semi-supervisé des étiqueteurs POS

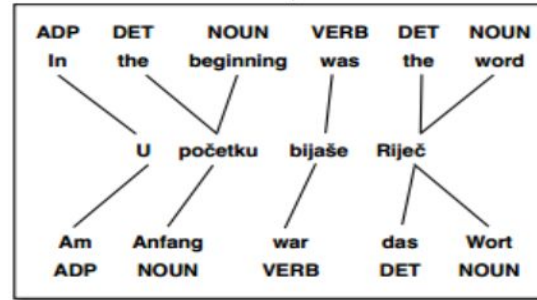
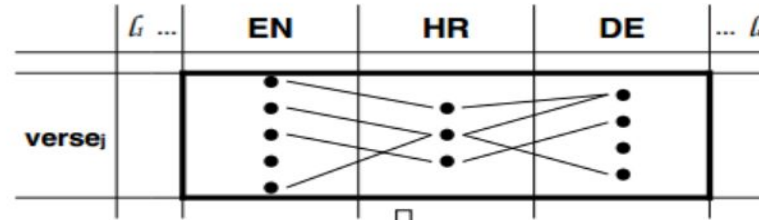
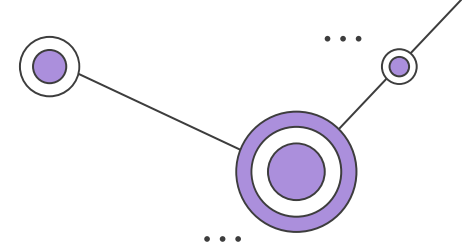
- Entraînement des étiqueteurs Part Of Speech pour chaque langue en utilisant à la fois les annotations projetées et les dictionnaires d'étiquettes.

03

## Evaluation

- Evaluation sur 25 langues (test existant).
- Evaluation sur 10 langues (dictionnaires d'étiquettes)
- Comparaison avec des baselines non supervisées en des étiqueteurs POS faiblement supervisées.

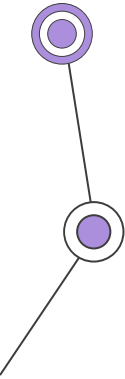




HR	EN	DE	...	voted	confidence
U	ADP	ADP	...	ADP	0.8667
početku	NOUN, DET	NOUN	...	NOUN	0.7448
bijaše	VERB	VERB	...	VERB	0.8560
Riječ	DET, NOUN	DET, NOUN	...	NOUN	0.6307

Figure 1: An illustration of our approach.

Matrice de  
projections  
d'annotations



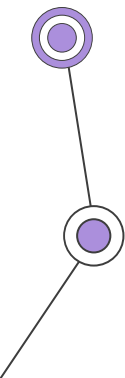
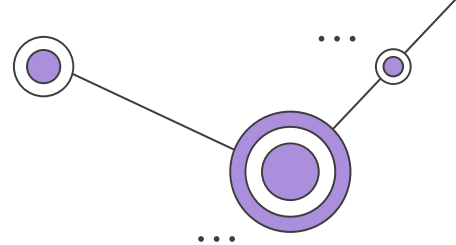
# Algorithmme

---

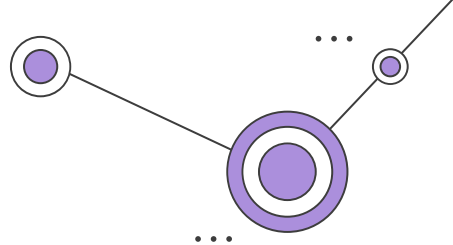
**Algorithm 1** Train  $n$  taggers with supervision for  $k$ 

---

- 1: Let  $M$  be a tensor with  $M(i, j, \cdot)$  the word-aligned token sequence in the  $j$ th verse of the Bible in language  $i$
  - 2: **for**  $i \leq k$  **do**
  - 3:   Train TNT tagger for  $l_i$  using manually annotated data
  - 4:   **for**  $j \leq m$  **do**
  - 5:     Obtain POS predictions for  $M(i, j, \cdot)$
  - 6:   **end for**
  - 7: **end for**
  - 8: **for**  $I \in \{0, 1\}$  **do**
  - 9:   **if**  $i > k, I = 1$  **then**
  - 10:     Train TNT tagger for  $l_i$  using projected annotations in  $M(i, \cdot, \cdot)$
  - 11:   **end if**
  - 12:   Populate  $M(i, \cdot, \cdot)$  by propagating tags across alignments
  - 13:   **for**  $i \leq n$  **do**
  - 14:     Use majority voting to obtain one tag per word
  - 15:     Obtain type-level tag dictionary from all the data
  - 16:     Train TNT/GAR tagger for  $l_i$  using projected annotations in  $M(i, \cdot, \cdot)$  and tag dictionary
  - 17:   **end for**
  - 18: **end for**
- 

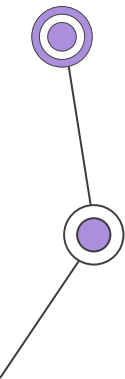


# Modèles



1. Etiqueteur POS **Tnt** (Trigrams'n'Tags)
2. Entraînement étiqueteur POS partiellement-supervisé avec l'approche Garrette et Baldrige (**GAR**)

Entraînement avec **k** et **n-1** langues , ce qui donne 4 étiqueteurs en total





# Résultats et évaluations

			UNSUPERVISED						UPPER BOUNDS			
			BASELINES		OUR SYSTEMS				WEAKLY SUP		SUPERVISED	
		OOV	BROWN	2HMM	TNT- <i>k</i> -SRC	TNT- <i>n</i> -1-SRC	GAR- <i>k</i> -SRC	GAR- <i>n</i> -1-SRC	DAS	LI	GAR	TNT
bul	YT	31.8	54.5	71.8	<b>78.0</b>	77.7	75.7	75.7	-	-	83.1	96.9
ces	YT	44.3	51.9	66.3	71.7	<b>73.3</b>	70.9	71.4	-	-	-	98.7
dan	YT	28.6	58.6	69.6	78.6	<b>79.0</b>	73.7	73.3	83.2	83.3	78.8	96.7
deu	YT	36.8	45.3	70.0	<b>80.5</b>	80.2	77.6	77.6	82.8	85.8	87.1	98.1
eng	YT	38.0	58.2	62.6	72.4	<b>73.0</b>	72.2	72.6	-	87.1	80.8	96.7
eus	NT	<u>64.6</u>	46.0	41.6	<b>63.4</b>	62.8	57.3	56.9	-	-	66.9	93.7
fra	YT	26.1	42.0	76.5	76.1	76.6	78.6	<b>80.2</b>	-	-	85.5	95.1
ell	YT	<u>63.7</u>	43.0	49.8	51.9	52.3	57.9	<b>59.0</b>	82.5	79.2	64.4	-
hin	Y	36.1	59.5	69.2	<b>70.9</b>	67.6	70.8	71.5	-	-	-	-
hrv	Y	34.7	52.8	65.6	<b>67.8</b>	67.1	67.2	66.7	-	-	-	-
hun	YT	41.2	45.9	57.4	70.0	70.4	71.3	<b>72.0</b>	-	-	77.9	95.6
isl	Y	19.7	42.6	65.9	<b>70.6</b>	69.0	68.7	68.3	-	-	-	-
ind	YT	29.4	52.6	73.1	76.6	<b>76.8</b>	74.9	76.0	-	-	87.1	95.1
ita	YT	24.0	45.1	78.3	76.5	76.9	78.5	<b>79.2</b>	86.8	86.5	83.5	95.8
plt	Y	35.0	48.9	44.3	56.4	56.6	62.0	<b>64.6</b>	-	-	-	-
mar	Y	33.0	<b>55.8</b>	45.8	52.0	52.9	52.8	52.3	-	-	-	-
nor	YT	27.5	56.1	73.0	<b>77.0</b>	76.7	75.4	76.0	-	-	84.3	97.7
pes	Y	33.6	57.9	<b>61.5</b>	59.3	59.6	59.1	60.8	-	-	-	-
pol	YT	36.4	52.2	68.7	<b>75.6</b>	75.1	70.8	74.0	-	-	-	95.7
por	YT	27.9	54.5	74.3	82.9	<b>83.8</b>	81.1	82.0	87.9	84.5	87.3	96.8
slv	Y	15.8	42.1	78.1	79.5	<b>80.5</b>	68.7	70.1	-	-	-	-
spa	YT	21.9	52.6	47.3	81.1	<b>82.6</b>	81.4	<b>82.6</b>	84.2	86.4	88.7	96.2
srp	Y	41.7	59.3	47.3	<b>69.6</b>	69.2	67.9	67.2	-	-	-	94.7
swe	YT	31.5	58.5	68.4	74.7	<b>75.2</b>	71.4	71.9	80.5	86.1	76.1	94.7
tur	YT	41.6	53.7	46.8	60.5	<b>61.3</b>	56.5	57.9	-	-	72.2	89.1
average	≤ 50		52.2	64.4	72.1	72.2	70.8	71.5				

Table 1: Results on 25 test languages. Y=entire Bible available. N=only New Testament available. T=manually annotated data available for training (but not used to obtain results for the language itself). Unsupervised baselines are evaluated using optimal 1:1 mappings.

# Base Line

					UNSUPERVISED						UPPER BOUNDS			
			BASELINES		OUR SYSTEMS				WEAKLY SUP		SUPERVISED			
OOV			BROWN	2HMM	TNT- <i>k</i> -SRC	TNT- <i>n</i> -1-SRC	GAR- <i>k</i> -SRC	GAR- <i>n</i> -1-SRC	DAS	LI	GAR	TNT		
bul	YT	31.8	54.5	71.8	<b>78.0</b>	77.7	75.7	75.7	-	-	83.1	96.9		
ces	YT	44.3	51.9	66.3	71.7	<b>73.3</b>	70.9	71.4	-	-	-	98.7		
dan	YT	28.6	58.6	69.6	78.6	<b>79.0</b>	73.7	73.3	83.2	83.3	78.8	96.7		
deu	YT	36.8	45.3	70.0	<b>80.5</b>	80.2	77.6	77.6	82.8	85.8	87.1	98.1		
eng	YT	38.0	58.2	62.6	72.4	<b>73.0</b>	72.2	72.6	-	87.1	80.8	96.7		
eus	NT	<u>64.6</u>	46.0	41.6	<b>63.4</b>	62.8	57.3	56.9	-	-	66.9	93.7		
fra	YT	26.1	42.0	76.5	76.1	76.6	78.6	<b>80.2</b>	-	-	85.5	95.1		
ell	YT	<u>63.7</u>	43.0	49.8	51.9	52.3	57.9	<b>59.0</b>	82.5	79.2	64.4	-		
hin	Y	36.1	59.5	69.2	<b>70.9</b>	67.6	70.8	71.5	-	-	-	-		
hrv	Y	34.7	52.8	65.6	<b>67.8</b>	67.1	67.2	66.7	-	-	-	-		
hun	YT	41.2	45.9	57.4	70.0	70.4	71.3	<b>72.0</b>	-	-	77.9	95.6		
isl	Y	19.7	42.6	65.9	<b>70.6</b>	69.0	68.7	68.3	-	-	-	-		
ind	YT	29.4	52.6	73.1	76.6	<b>76.8</b>	74.9	76.0	-	-	87.1	95.1		
ita	YT	24.0	45.1	78.3	76.5	76.9	78.5	<b>79.2</b>	86.8	86.5	83.5	95.8		
plt	Y	35.0	48.9	44.3	56.4	56.6	62.0	<b>64.6</b>	-	-	-	-		
mar	Y	33.0	<b>55.8</b>	45.8	52.0	52.9	52.8	52.3	-	-	-	-		
nor	YT	27.5	56.1	73.0	<b>77.0</b>	76.7	75.4	76.0	-	-	84.3	97.7		
pes	Y	33.6	57.9	<b>61.5</b>	59.3	59.6	59.1	60.8	-	-	-	-		
pol	YT	36.4	52.2	68.7	<b>75.6</b>	75.1	70.8	74.0	-	-	-	95.7		
por	YT	27.9	54.5	74.3	82.9	<b>83.8</b>	81.1	82.0	87.9	84.5	87.3	96.8		
slv	Y	15.8	42.1	78.1	79.5	<b>80.5</b>	68.7	70.1	-	-	-	-		
spa	YT	21.9	52.6	47.3	81.1	81.4	<b>82.6</b>	<b>82.6</b>	84.2	86.4	88.7	96.2		
srp	Y	41.7	59.3	47.3	<b>69.6</b>	69.2	67.9	67.2	-	-	-	94.7		
swe	YT	31.5	58.5	68.4	74.7	<b>75.2</b>	71.4	71.9	80.5	86.1	76.1	94.7		
tur	YT	41.6	53.7	46.8	60.5	<b>61.3</b>	56.5	57.9	-	-	72.2	89.1		
average	≤ 50		52.2	64.4	72.1	72.2	70.8	71.5						

Table 1: Results on 25 test languages. Y=entire Bible available. N=only New Testament available. T=manually annotated data available for training (but not used to obtain results for the language itself). Unsupervised baselines are evaluated using optimal 1:1 mappings.

# TnT

					UNSUPERVISED		UPPER BOUNDS					
			BASELINES		OUR SYSTEMS				WEAKLY SUP		SUPERVISED	
OOV			BROWN	2HMM	TnT- <i>k</i> -SRC	TnT- <i>n</i> -1-SRC	GAR- <i>k</i> -SRC	GAR- <i>n</i> -1-SRC	DAS	LI	GAR	TnT
bul	YT	31.8	54.5	71.8	<b>78.0</b>	77.7	75.7	75.7	-	-	83.1	96.9
ces	YT	44.3	51.9	66.3	71.7	<b>73.3</b>	70.9	71.4	-	-	-	98.7
dan	YT	28.6	58.6	69.6	78.6	<b>79.0</b>	73.7	73.3	83.2	83.3	78.8	96.7
deu	YT	36.8	45.3	70.0	<b>80.5</b>	80.2	77.6	77.6	82.8	85.8	87.1	98.1
eng	YT	38.0	58.2	62.6	72.4	<b>73.0</b>	72.2	72.6	-	87.1	80.8	96.7
eus	NT	<u>64.6</u>	46.0	41.6	<b>63.4</b>	62.8	57.3	56.9	-	-	66.9	93.7
fra	YT	26.1	42.0	76.5	76.1	76.6	78.6	<b>80.2</b>	-	-	85.5	95.1
ell	YT	<u>63.7</u>	43.0	49.8	51.9	52.3	57.9	<b>59.0</b>	82.5	79.2	64.4	-
hin	Y	36.1	59.5	69.2	<b>70.9</b>	67.6	70.8	71.5	-	-	-	-
hrv	Y	34.7	52.8	65.6	<b>67.8</b>	67.1	67.2	66.7	-	-	-	-
hun	YT	41.2	45.9	57.4	70.0	70.4	71.3	<b>72.0</b>	-	-	77.9	95.6
isl	Y	19.7	42.6	65.9	<b>70.6</b>	69.0	68.7	68.3	-	-	-	-
ind	YT	29.4	52.6	73.1	76.6	<b>76.8</b>	74.9	76.0	-	-	87.1	95.1
ita	YT	24.0	45.1	78.3	76.5	76.9	78.5	<b>79.2</b>	86.8	86.5	83.5	95.8
plt	Y	35.0	48.9	44.3	56.4	56.6	62.0	<b>64.6</b>	-	-	-	-
mar	Y	33.0	<b>55.8</b>	45.8	52.0	52.9	52.8	52.3	-	-	-	-
nor	YT	27.5	56.1	73.0	<b>77.0</b>	76.7	75.4	76.0	-	-	84.3	97.7
pes	Y	33.6	57.9	<b>61.5</b>	59.3	59.6	59.1	60.8	-	-	-	-
pol	YT	36.4	52.2	68.7	<b>75.6</b>	75.1	70.8	74.0	-	-	-	95.7
por	YT	27.9	54.5	74.3	82.9	<b>83.8</b>	81.1	82.0	87.9	84.5	87.3	96.8
slv	Y	15.8	42.1	78.1	79.5	<b>80.5</b>	68.7	70.1	-	-	-	-
spa	YT	21.9	52.6	47.3	81.1	81.4	<b>82.6</b>	<b>82.6</b>	84.2	86.4	88.7	96.2
srp	Y	41.7	59.3	47.3	<b>69.6</b>	69.2	67.9	67.2	-	-	-	94.7
swe	YT	31.5	58.5	68.4	74.7	<b>75.2</b>	71.4	71.9	80.5	86.1	76.1	94.7
tur	YT	41.6	53.7	46.8	60.5	<b>61.3</b>	56.5	57.9	-	-	72.2	89.1
average	≤ 50		52.2	64.4	72.1	72.2	70.8	71.5				

Table 1: Results on 25 test languages. Y=entire Bible available. N=only New Testament available. T=manually annotated data available for training (but not used to obtain results for the language itself). Unsupervised baselines are evaluated using optimal 1:1 mappings.

# GAR

			UNSUPERVISED						UPPER BOUNDS			
			BASELINES		OUR SYSTEMS				WEAKLY SUP		SUPERVISED	
		OOV	BROWN	2HMM	TNT- <i>k</i> -SRC	TNT- <i>n</i> -1-SRC	GAR- <i>k</i> -SRC	GAR- <i>n</i> -1-SRC	DAS	LI	GAR	TNT
bul	YT	31.8	54.5	71.8	<b>78.0</b>	77.7	75.7	75.7	-	-	83.1	96.9
ces	YT	44.3	51.9	66.3	71.7	<b>73.3</b>	70.9	71.4	-	-	-	98.7
dan	YT	28.6	58.6	69.6	78.6	<b>79.0</b>	73.7	73.3	83.2	83.3	78.8	96.7
deu	YT	36.8	45.3	70.0	<b>80.5</b>	80.2	77.6	77.6	82.8	85.8	87.1	98.1
eng	YT	38.0	58.2	62.6	72.4	<b>73.0</b>	72.2	72.6	-	87.1	80.8	96.7
eus	NT	<u>64.6</u>	46.0	41.6	<b>63.4</b>	62.8	57.3	56.9	-	-	66.9	93.7
fra	YT	26.1	42.0	76.5	76.1	76.6	78.6	<b>80.2</b>	-	-	85.5	95.1
ell	YT	<u>63.7</u>	43.0	49.8	51.9	52.3	57.9	<b>59.0</b>	82.5	79.2	64.4	-
hin	Y	36.1	59.5	69.2	<b>70.9</b>	67.6	70.8	71.5	-	-	-	-
hrv	Y	34.7	52.8	65.6	<b>67.8</b>	67.1	67.2	66.7	-	-	-	-
hun	YT	41.2	45.9	57.4	70.0	70.4	71.3	<b>72.0</b>	-	-	77.9	95.6
isl	Y	19.7	42.6	65.9	<b>70.6</b>	69.0	68.7	68.3	-	-	-	-
ind	YT	29.4	52.6	73.1	76.6	<b>76.8</b>	74.9	76.0	-	-	87.1	95.1
ita	YT	24.0	45.1	78.3	76.5	76.9	78.5	<b>79.2</b>	86.8	86.5	83.5	95.8
plt	Y	35.0	48.9	44.3	56.4	56.6	62.0	<b>64.6</b>	-	-	-	-
mar	Y	33.0	<b>55.8</b>	45.8	52.0	52.9	52.8	52.3	-	-	-	-
nor	YT	27.5	56.1	73.0	<b>77.0</b>	76.7	75.4	76.0	-	-	84.3	97.7
pes	Y	33.6	57.9	<b>61.5</b>	59.3	59.6	59.1	60.8	-	-	-	-
pol	YT	36.4	52.2	68.7	<b>75.6</b>	75.1	70.8	74.0	-	-	-	95.7
por	YT	27.9	54.5	74.3	82.9	<b>83.8</b>	81.1	82.0	87.9	84.5	87.3	96.8
slv	Y	15.8	42.1	78.1	79.5	<b>80.5</b>	68.7	70.1	-	-	-	-
spa	YT	21.9	52.6	47.3	81.1	81.4	<b>82.6</b>	<b>82.6</b>	84.2	86.4	88.7	96.2
srp	Y	41.7	59.3	47.3	<b>69.6</b>	69.2	67.9	67.2	-	-	-	94.7
swe	YT	31.5	58.5	68.4	74.7	<b>75.2</b>	71.4	71.9	80.5	86.1	76.1	94.7
tur	YT	41.6	53.7	46.8	60.5	<b>61.3</b>	56.5	57.9	-	-	72.2	89.1
average	≤ 50		52.2	64.4	72.1	72.2	70.8	71.5				

Table 1: Results on 25 test languages. Y=entire Bible available. N=only New Testament available. T=manually annotated data available for training (but not used to obtain results for the language itself). Unsupervised baselines are evaluated using optimal 1:1 mappings.



# Comparaison Wiktionnaire

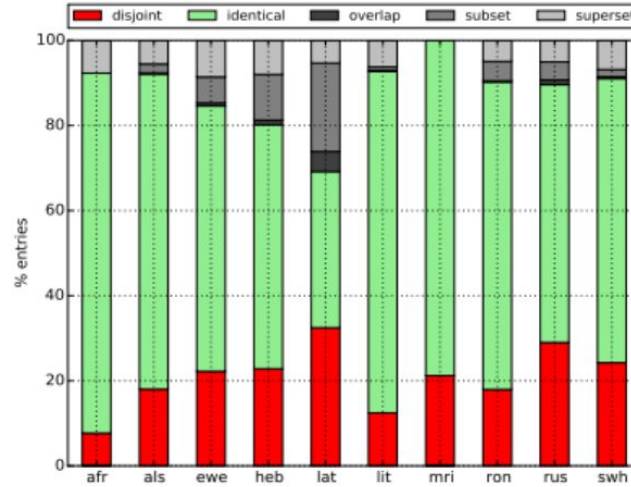


Figure 2: Type-level in-vocabulary tag errors as the percentage of word types assigned a set of tags that is disjoint, identical to, overlaps, is a subset, or is a superset of the Wiktionary tags.

# Calcul d'exactitude (*accuracy*) au niveau des étiquettes ...

## Wiktionnaire

Afrikaans, Lituanien, Russe : 80-90%  
Hébreu, Roumain , Swahili : 50-70%  
Latin, Maori, Albanien, Ewe: entre 35 et 50%

# Objectif atteint !

La performance du modèle **20-30% meilleure** que les nouvelles techniques développées pour les POS taggers issues des traductions de la Bible.

Le modèle se montre compétitifs par rapport aux approches développées pour des langues plus dotées en ressources.

De plus, le modèle a été rendu public pour les 100 langues étudiés et le mapping de Petrov et al. (2011) a été étendu à six nouvelles langues (hindi, croate, islandais, norvégien, persan et serbe.)

# Est ce qu'on peut s'appropriier le système dans un cadre différent ?

- Ce corpus appartient un style particulier biblique, littéraire
- À tester sur d'autres corpus parallèles (médical, journalistique)
  - risque d'avoir des moins bon résultats avec d'autres corpus





# Refaire la même expérience permettra-t-il d'obtenir des résultats similaires ?

L'utilisation du **même corpus biblique** avec une **méthodologie identique** produira des **résultats similaires**. Cependant, l'exploration de **méthodologies** plus **innovantes** et **adaptées** pourrait potentiellement conduire à des **améliorations**.

Dans ce cas, en conservant la **même base de données**, l'introduction de **techniques novatrices** pourrait permettre **d'optimiser** la **performance** du **système**, offrant ainsi une avenue intéressante pour des résultats plus prometteurs.

Exemple :

- Enrichir les Données d'Entraînement
- Utiliser des Ressources Linguistiques Complémentaires (dictionnaires linguistiques, lexique.)
- Optimiser la Méthode de Projection d'Annotation.





Dan Garette & Jason Baldrige

**“Learning a  
Part-of-Speech Tagger  
from Two Hours of  
Annotation”**



# Les particularités de cette expérience sont :

01

## Les langues traitées :

Les langues peu dotées, non indo-européennes

02

## L'annotation :

Une campagne d'annotation manuelle réalisée en 2 heures !

03

## Les données :

2 types de données :

- Dictionnaire de mots
- Des phrases étiquetées



# Les données :



## Les ressources en langues peu dotées

- Le malgache (les articles journalistiques trouvés sur les sites);
- Le kinyarwanda (transcriptions des survivants des génocides);
- + l'anglais (Penn Treebank) pour la vérification

## On traite 2 types de données supervisées

- Sur les mots
- Sur les phrases

-> Plus concrètement : les tags de dictionnaires et les phrases labellisées

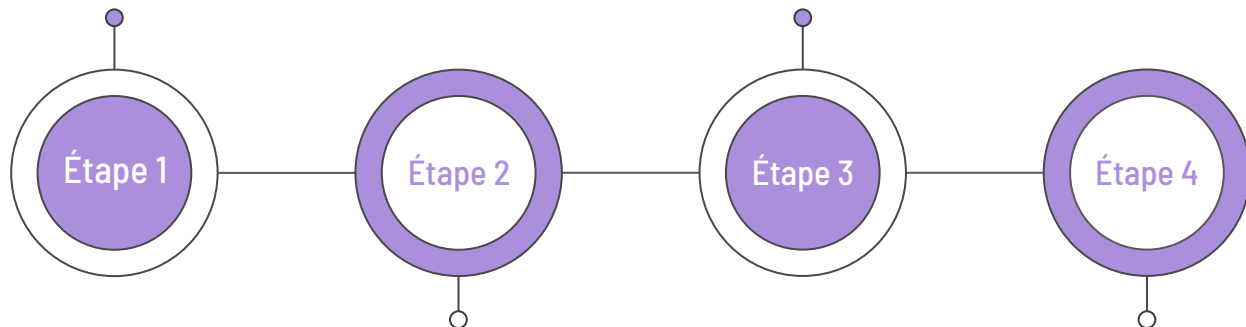
## 4 sets utilisés (par langue)

1. Corpus train à annoter (1000-2000 tokens/types annotés par non natifs)
2. Un corpus train brut (100 000 tokens par langue)
3. Dev
4. Test

# Méthodologie

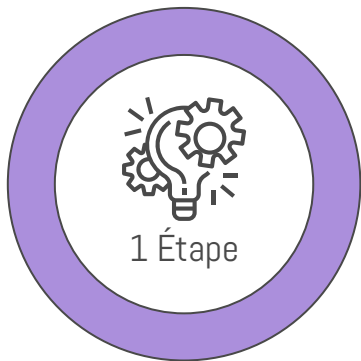
Extension du dictionnaire  
de mots-étiquettes  
(Tag dictionary expansion)

Expectation Minimization  
HMM training



Minimisation du modèle  
pondéré  
(Weighted model minimization)

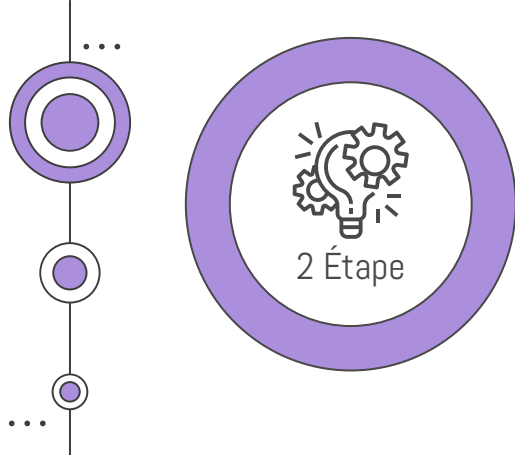
Maximal Entropy model  
training



# Tag Dictionary Expansion

- **Petite quantité de ressources:** un dictionnaire tel quel issu de l'annotation ne suffit pas pour entraîner HMM.
- **Méthodologie intéressante :** Pour élargir ce dictionnaire, un "label propagation" est utilisé. C'est une technique sous forme de graphe pour répartir les étiquettes entre les éléments apparentés. Ils ont également étudié l'efficacité de l'utilisation d'un dictionnaire externe dans le graphe LP (Wikitionary).
- **Défis :** la qualité des données annotées automatiquement.





# Weighted model minimization

- **Qualité des ressources** : après avoir élargie automatiquement un dictionnaire, il y a plus de bruit.
- **But** : La procédure pour trouver un modèle qui garde les bigrammes d'étiquettes qui ont constamment des poids élevés dans l'ensemble du corpus.
- **Résultat** : le chemin de tags avec les poids les plus élevés à travers chaque phrase du corpus d'entraînement => une base pour initialiser l'étape suivante

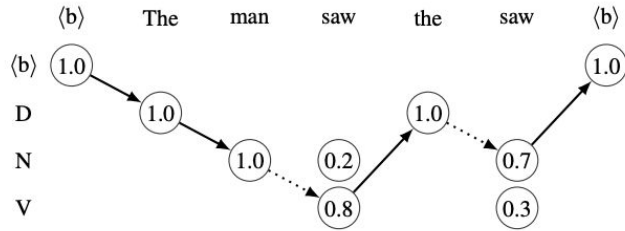


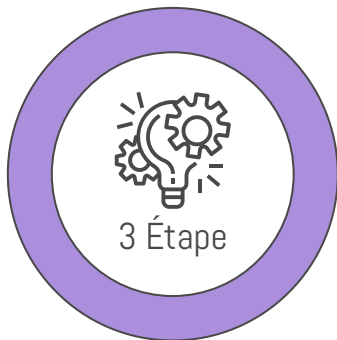
Figure 2: Weighted, greedy model minimization graph showing a potential state between the stages of the tag bigram choosing algorithm. Solid edges: selected bigrams. Dotted edges: holes in the path.

# Expectation minimization HMM training

- **Les données** : utilise le corpus automatiquement annoté dans l'étape précédente pour entraîner le HMM.
- **Le but** : Le HMM entraîné n'est pas utilisé directement dans l'étiquetage. Il est concaténé avec le token-supervised corpus (si disponible) pour entraîner un Markov à entropie maximale.

## MaxEnt Markov model training

- Le MEMM exploite les "subword features" et produit généralement 1 à 2 % de meilleurs résultats qu'un HMM





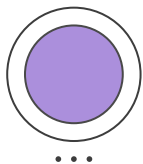
# Les résultats

- La performance s'améliore avec les techniques de LP (surtout) et minimization par rapport à l'entraînement EM-HMM.
- Différences entre token et type supervised annotations.
- Il semble préférable de collecter un dictionnaire d'étiquettes plutôt que des phrases étiquetées.

Human Annotations	0. No EM			1. EM only			2. With LP			3. LP+min			4. LP(ed)+min		
Initial data	T	K	U	T	K	U	T	K	U	T	K	U	T	K	U
KIN tokens A	72	90	58	55	82	32	71	86	58	71	86	58	71	86	58
KIN types A				63	77	32	78	83	69	79	83	70	79	83	70
MLG tokens B	74	89	49	68	87	39	74	89	49	74	89	49	76	90	53
MLG types B				71	87	46	72	81	57	74	86	56	76	86	60
ENG tokens A	63	83	38	62	83	37	72	85	55	72	85	55	72	85	56
ENG types A				66	76	37	75	81	56	76	83	56	74	81	55
ENG tokens B	70	87	44	70	87	43	78	90	60	78	90	60	78	89	61
ENG types B				69	83	38	75	82	61	78	85	61	78	86	61

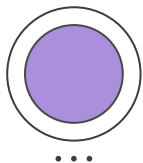
Table 2: Experimental results. Three languages are shown: Kinyarwanda (KIN), Malagasy (MLG), and English (ENG). The letters A and B refer to the annotator. LP(ed) refers to label propagation including nodes from an external dictionary. Each result given as percentages for Total (T), Known (K), and Unknown (U).

# Un retour critique...



## Objectif du projet

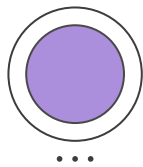
L'impression prédominante est que le projet avait pour principal objectif de tester une méthode plutôt que de générer un résultat significatif.



## Fiabilité des données

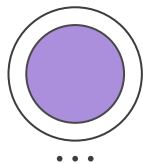
Les données de départ ne sont pas fiables.

Annotateurs non natifs et utilisation de Wiktionary comme dictionnaire.



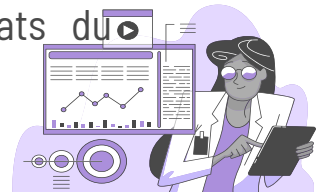
## Wiktionary en chiffres :

Les étiquettes de Wiktionary étaient valides à seulement 80-90% pour trois langues. Pour des langues moins courantes (l'albanais, l'ewe, etc.), la validité chutait à seulement 35-50%.

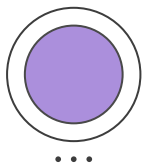


## Scénario

La situation de 2 heures sans révision d'annotation peut entraîner un niveau élevé de bruit dans les résultats du projet.



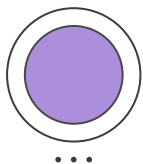
# Est-ce que nous pouvons nous approprier ce système dans un cadre différent ?



## Cela peut être productif et efficace.

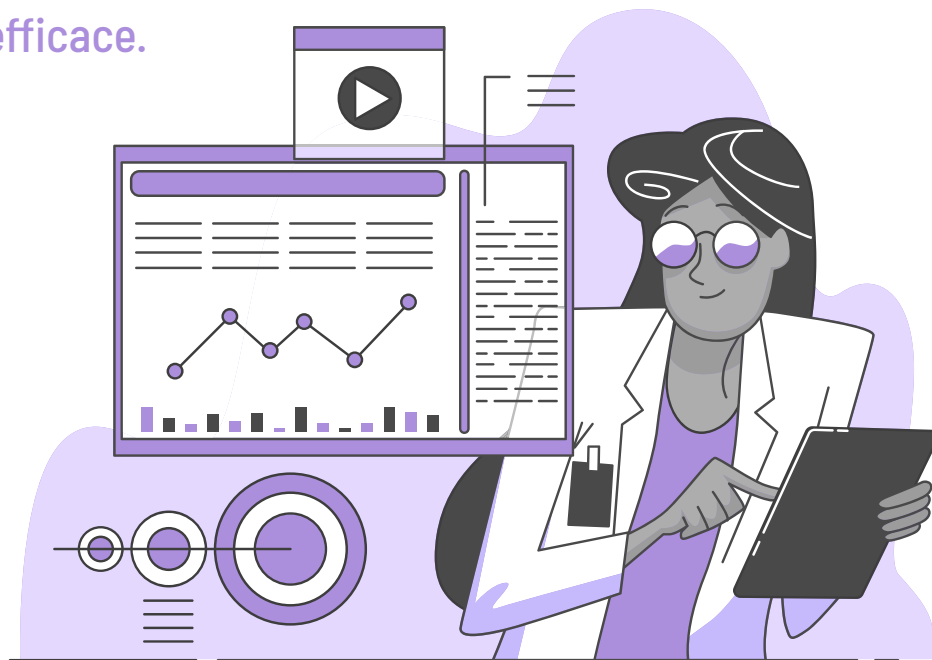
Surtout en s'assurant de pouvoir avoir des données d'entraînement de qualité.

(Des annotateurs natifs et bien formés, des sources fiables etc.)

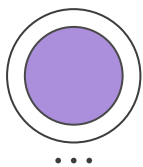


## La disponibilité

Pas de traitement d'une œuvre donnée et pas de restrictions techniques majeures telles que la nécessité des alignements.

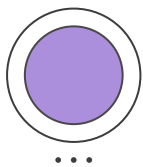


# Est-ce que nous pouvons répéter l'expérience et obtenir les mêmes résultats ?



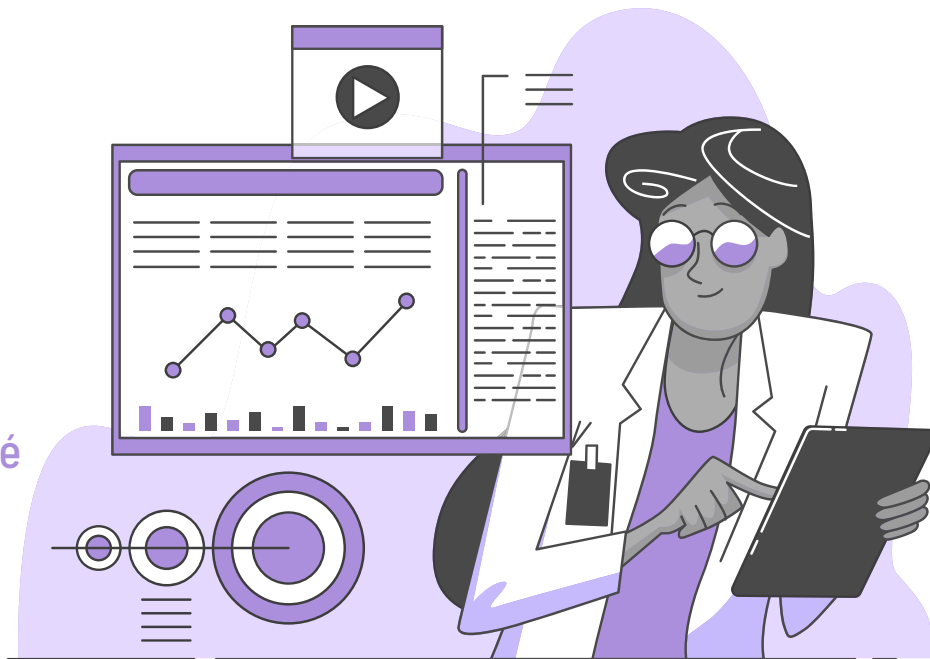
## Pas forcément...

Il y a peu de chance que deux annotateurs qui annotent manuellement reproduisent exactement les mêmes étiquettes, d'où peuvent ressortir des différences dans les résultats.



## Résultat dépendant de l'unité

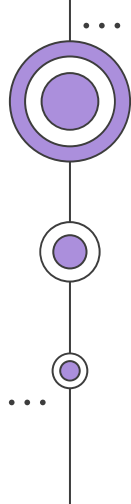
L'expérience sur l'anglais a démontré que la vitesse et la qualité de leur annotations influencent les résultats



# Que peut-on dire à propos des deux articles ?

Apprentissage d'étiqueteur de parties du discours (POS) sur des langues peu dotés de ressources.





Article n°1

Article n°2

Annotation

Semi-supervisé

Supervisé et chronométré

Entrainement

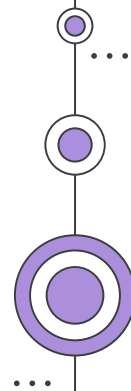
TnT, GAR

LP, HMM, EMHMM, MEMM

Résultats

De très bons résultats ! 😊

Les résultats ok ... 😞



# Bibliographie

- Željko Agić, Dirk Hovy, and Anders Søgaard. 2015. If all you have is a bit of the Bible: Learning POS taggers for truly low-resource languages. In *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 268–272, Beijing, China. Association for Computational Linguistics.
- Dan Garrette and Jason Baldridge. 2013. Learning a Part-of-Speech Tagger from Two Hours of Annotation. In *Proceedings of the 2013 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 138–147, Atlanta, Georgia. Association for Computational Linguistics.

Merci !



Des questions ?

