

Projet PSTALN

Prédiction des *morph*

Cléa Han, Yanis Labeyrie et Adrien Zabban

15 janvier 2024

Le but: prédire les *morphy*

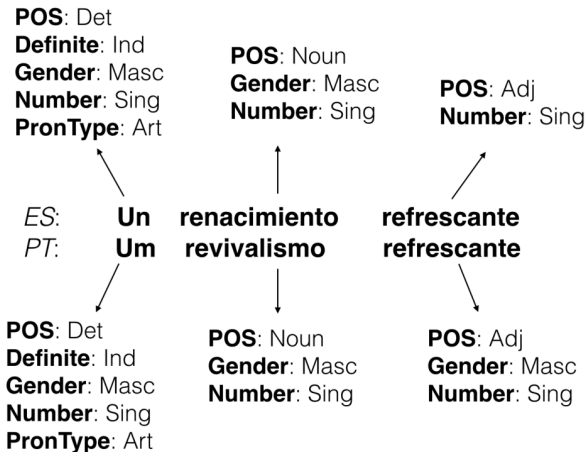


Figure: Tag morphologique d'une phrase portugaise et sa traduction en espagnol.

Nous avons utilisé le dataset Universal Dependencies 2.13.

Le dataset en français contient:

- 47498 phrases
- 849476 mots
- 76048 mots uniques

On a recensé:

- 19 classes *pos.*
- 28 classes *morphy*, avec un nombre de possibilités entre 2 et 13

Les Données: La gestion des mots inconnus

Modèle *GET_POS*

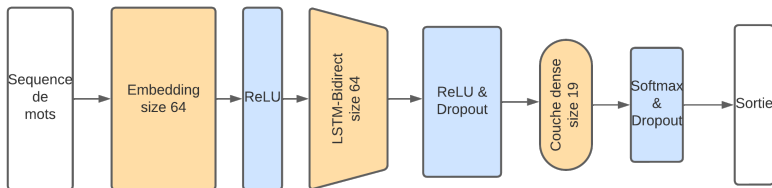


Figure: Modèle *GET_POS*

Modèle *SUPERTAG*

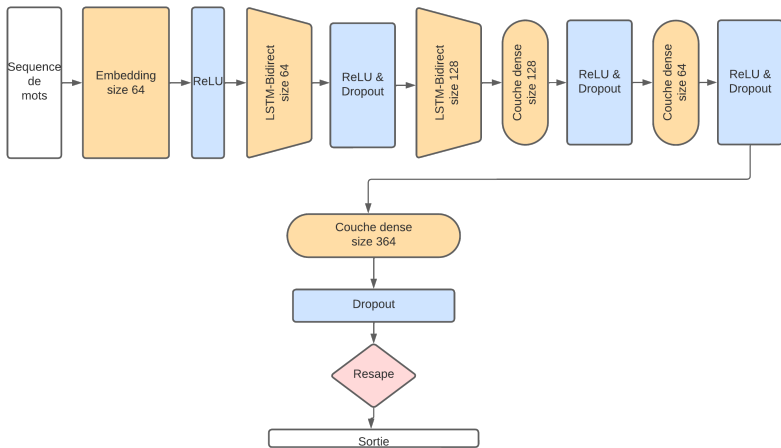


Figure: Modèle *SUPERTAG*

Modèle *SEPARATE*

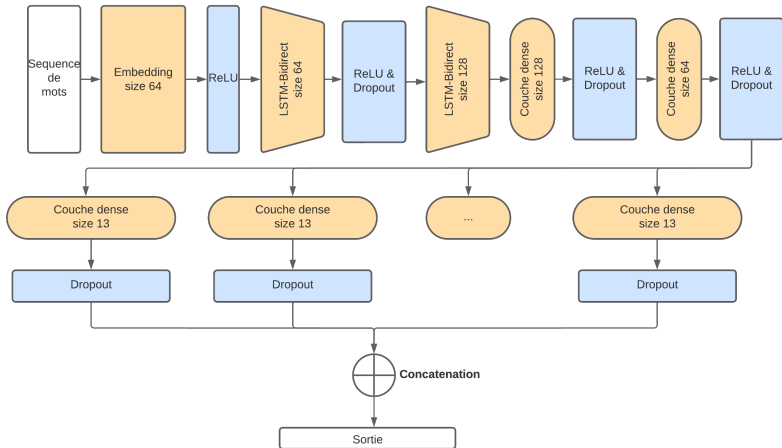


Figure: Modèle *SEPARATE*.

Modèle *FUSION*

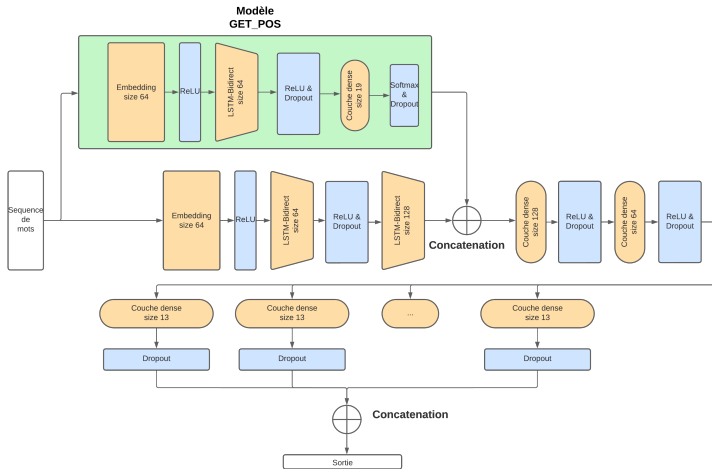


Figure: Modèle *FUSION*

Les loss:

- crossentropy pour le *pos*
- moyenne de la crossentropy sur les 28 classes pour le *morphy*

Les métriques:

- accuracy micro
- accuracy macro (pour le *pos*)
- allgood (pour le *morphy*)

Résultats: *GET_POS*

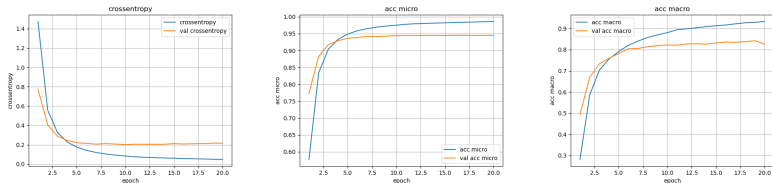


Figure: Entraînement du modèle *GET_POS*.

Nom du modèle	crossentropy	accuracy micro	accuracy macro
<i>GET_POS</i>	0.204	0.944	0.816

Table: Résultats du modèle *GET_POS* sur la base de données de teste.

Résultats des modèles *SUPERTAG* et *SEPARATE*

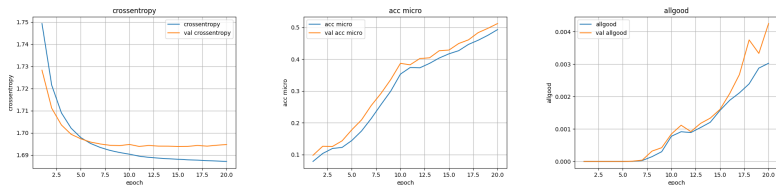


Figure: Entraînement du modèle *SUPERTAG*.

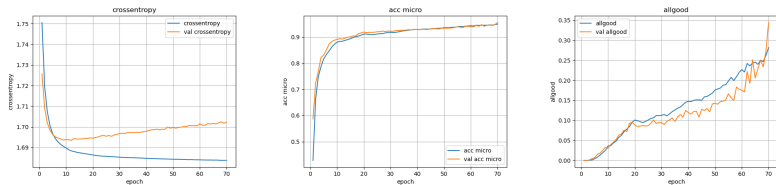


Figure: Entraînement du modèle *SEPARATE*.

Résultats du modèle *FUSION* et résultats de teste

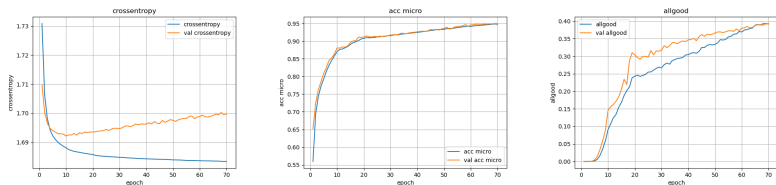


Figure: Entraînement du modèle *FUSION*.

Nom du modèle	crossentropy	accuracy micro	all good
<i>BASELINE</i>	-	0.980	0.791
<i>SUPERTAG</i>	1.700	0.436	0.002
<i>SEPARATE</i>	1.70	0.893	0.046
<i>FUSION</i>	1.698	0.884	0.154

Table: Résultats de test sur la prédiction des *morphy*

