

# Détection d'interactions neuronales d'un identifiant de langue



Baptiste Blouin

Jeudi 24 janvier 2019

# Sommaire

- ❏ Cadre de l'étude
- ❏ Expériences
- ❏ Résultats

# Typological Features for Multilingual Delexicalised Dependency Parsing

- ❑ Modèle de transitions :
  - ❑ Arc-Eager
  - ❑ Configuration (  $\sigma$  ,  $\beta$  , A )
- ❑ Perceptron multicouches ( MLP )

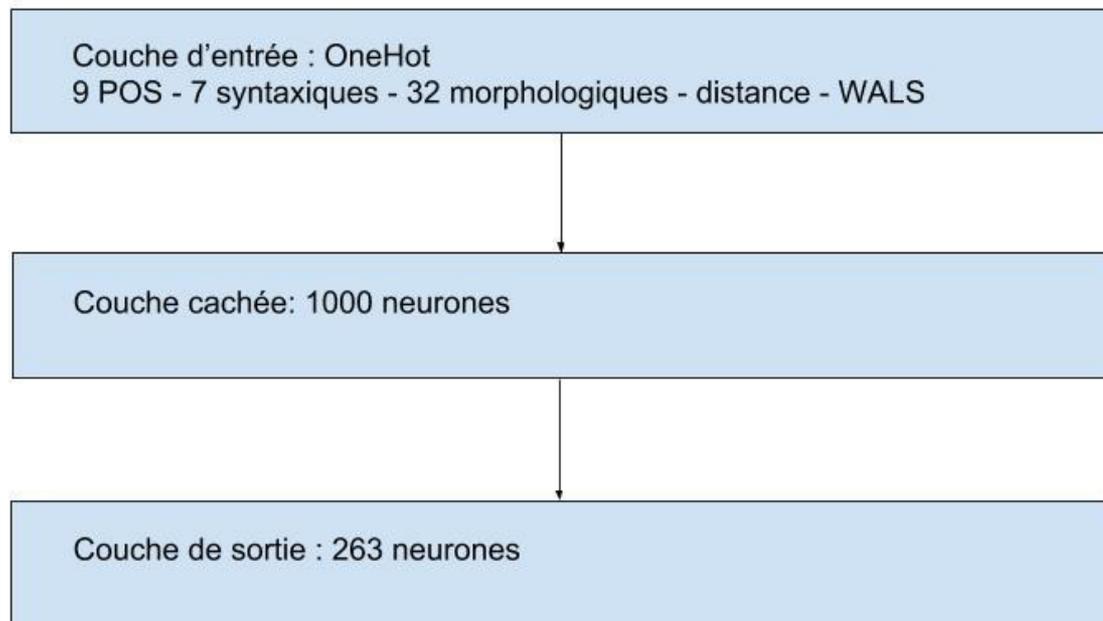
## Universal Dependencies ( UD )

- ❑ 40 langues utilisées
- ❑ 529 mots pour le Kazakh / 1.842.867 mots pour le tchèque
- ❑ Parties de discours ( POS )
- ❑ Autres caractéristiques

## World Atlas of Language Structures ( WALS )

- ❑ Phonologique, grammatical et lexical
- ❑ Décrit 2676 langues avec un ensemble de 192 features
- ❑ 87A: “Order of Adjective and Noun”

## Parser multilingue



## Raison de la recherche

- ❑ Amélioration de 4% avec l'ajout de WALS
- ❑ Amélioration de plus d'1% par rapport à un identifiant de langue
- ❑ WALS permet d'identifier la langue ? Ou juste de rajouter des features ?

# Expériences

## Recherche d'interactions potentielles

- ❑ Mise en correspondance UD et WALS
  - ❑ 87A: “Order of Adjective and Noun”
    - ❑ Adjectif - Nom pour l’anglais
    - ❑ Nom - Adjectif pour le français
  
- ❑ Recherche d’interactions inattendues
  - ❑ 87A: “Order of Adjective and Noun”
    - ❑ Verbe - Déterminant pour l’anglais

## Première approche

- ❑ Calcul de la moyenne des poids pour chaque features
- ❑ Recherche de correspondances entre les features

## Détection d'interactions

Neural Interaction Detection ( Tsang et al., 2018 )

- ❑ Classement d'interactions
- ❑ Recherche de coupure sur le classement
- ❑ Calcul de l'influence

$$\mathbf{z}^{(1)} = |\mathbf{w}^y|^\top \left| \mathbf{W}^{(L)} \right| \cdot \left| \mathbf{W}^{(L-1)} \right| \dots \left| \mathbf{W}^{(2)} \right|,$$
$$\omega_i(\mathcal{I}) = z_i^{(1)} \mu \left( \left| \mathbf{W}_{i,\mathcal{I}}^{(1)} \right| \right),$$

# Résultats

## Recherche d'interactions globales

 POS et WAL5

| Interactions   | Scores d'interactions |
|--|-----------------------|
| s.2.POS_19_PUNCT_., b.0.POS_29_PUNCT_  | 7.374025334463619     |
| b.0.83A_38_UNDEF, b.0.83A_37_UNDEF   | 6.889302434998262     |
| b.1.POS_70_PUNCT_>>, b.0.83A_38_UNDEF, b.0.83A_37_UNDEF  | 6.63804356125806      |
| s.2.POS_19_PUNCT_., b.0.POS_29_PUNCT_ , b.0.85A_18_UNDEF   | 6.617705454104581     |
| s.2.POS_19_PUNCT_., b.0.POS_29_PUNCT_ , b.1.POS_91_PUNCT_etcètera, b.0.85A_18_UNDEF                      | 6.545940486253886     |
| b.0.81A_39_UNDEF, b.0.81A_6_0  | 6.466659292681049     |
| s.2.POS_19_PUNCT_., b.0.POS_29_PUNCT_ , b.1.POS_125_PUNCT_<, b.1.POS_91_PUNCT_etcètera, b.0.85A_18_UNDEF | 6.442695914688982     |

## Recherche d'interactions centrées

### Interactions potentielles

b.0.87A\_2,b.0.POS\_ADJ,b.1.POS\_NOUN#English

Score :

[((1, 2), 0.5586156289748145), ((2, 3), 0.11473158782011751), ((1, 2, 3), 0.0786398441152464)

### Interactions inattendues

b.0.87A\_2,b.0.POS\_PRON,b.1.POS\_VERB#English

Score :

[((2, 3), 0.7662461057301234), ((1, 2, 3), 0.3068150378537299)]

## Pour aller plus loin

- ❑ Adapter l'ensemble d'apprentissage
- ❑ Tester sur l'ensemble du réseau
- ❑ Tester plus d'interactions

## Conclusion

- ❑ Utilité pour les langues particulières
- ❑ Rajout de features pour les langues communes

Merci pour votre attention !