

# Identification d'expressions verbales Verbal MWE identification with a neural transition-based parser

Hazem Al Saied(1), Marie Candito (2) et Mathieu Constant (1)  
(1)Université de Lorraine & CNRS, ATILF  
(2)Université Paris Diderot & CNRS, LLF

13 juin 2019



## Point de départ

- Données PARSEME multilingues annotées en expressions verbales (Savary et al. 2017, Ramisch et al. 2018)
- Système par transitions d'identification des expressions verbales (Al Saied et al. 2017)
  - Approche supervisée (closed track)
  - Modèle linéaire (SVM)
  - Données en entrée (lemmes des mots, étiquetage morphosyntaxique, analyse syntaxique en dépendances)
  - Jeu de traits optimisé manuellement par langue

## Questions

- Passage à une architecture neuronale sans annotation syntaxique en entrée ?
- Quelle procédure multilingue de recherche de la meilleure configuration d'hyperparamètres ?

## Quelle architecture ?

- **Simple perceptron multi-couches**
- vs. architecture à base de réseau récurrent sur les mots (ex. Kiperwasser et Goldberg 2016)

## Expériences préliminaires catastrophiques

- Jeu de données déséquilibré (rareté des expressions verbales)  
→ ré-échantillonnage
- Besoin d'une méthode rigoureuse de sélection des hyperparamètres

# Ensemble d'hyperparamètres

Type	Hyperparameter	Range or set	BoR <sub>c</sub>	BoR <sub>o</sub>	TB
Embedding and initialisation	Use $B_2$	{True, False}	True	True	True
	Use $B_{-1}$	{True, False}	True	False	True
	Lemmatization	{True, False}	True	True	True
	Token dimension	[100, 600]	157	300	300
	POS dimension	[15, 150]	147	132	35
	Pre-trained	{True, False}	False	True	True   False
	Trainable	{True, False}	True	True	True
	Averaging	{True, False}	False	True	True
	Vocabulary	{Compact, Exhaustive}	True	False	True
Dev set	Unit number	[25, 600]	85	56	75
	Dropout	{.1, .2, ... .6}	0.3	0.1	0.4
Sampling	Focused / Frequency threshold	{True, False} / {5, 10, ... 30}	False / -	False / -	False / -
	Over loss / Loss coefficient	{True, False} / [1, 40]	False / 1	False / 1	False / 1
Train	Learning rate	[.01, .2]	0.017	0.095	0.03
	Batch size	{16, 32, 48, 64, 128}	128	16	48

# Recherche des hyperparamètres (simplifié)

- Choix de trois langues pilotes (BG, PT, TR),
- Réduction des données pilote à une taille moyenne
- Recherche aléatoire des hyperparamètres (Bergstra et Bengio, 2012)
- Choix de la meilleure configuration par tendances individuelles des hyperparamètres
- Application de la meilleure configuration sur toutes les langues

- Résultats finaux sur les données PARSEME-FR 1.1 : neuronal (F1 = 61.3) vs. linéaire (59.3)
- Comparaisons : meilleur système de la Shared Task (54.0), score du meilleur système de la Shared Task par langue (56.9)
- Recherche aléatoire (61.2 sur dev) vs. tendances (63.5)
- Score quasi-nul sur les expressions inconnues

# Conclusions

- Passage au neuronal performant, mais à quel prix !
- Modèle linéaire bien plus simple à optimiser, pour des performances proches du modèle neuronal
- Pouvoir de généralisation nul
- Perspectives : attention is all you need, multitasking, contextualized embeddings