

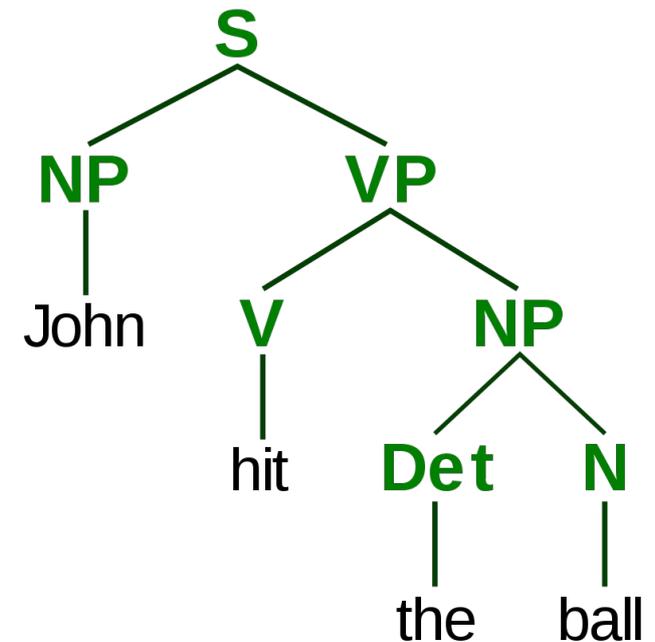
Présentation UE de Communication

Article : Typological Features for
Multilingual Delexicalised
Dependency Parsing



L'analyse syntaxique

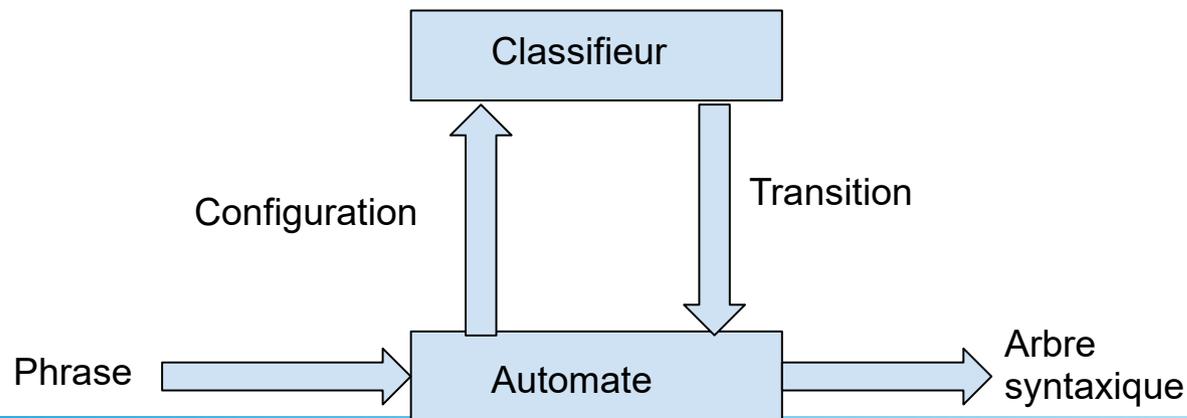
- Compréhension de la structure d'une phrase
- Appartenance à un langage ?
- Liens entre les mots ?
- Arbre syntaxique



L'analyseur syntaxique

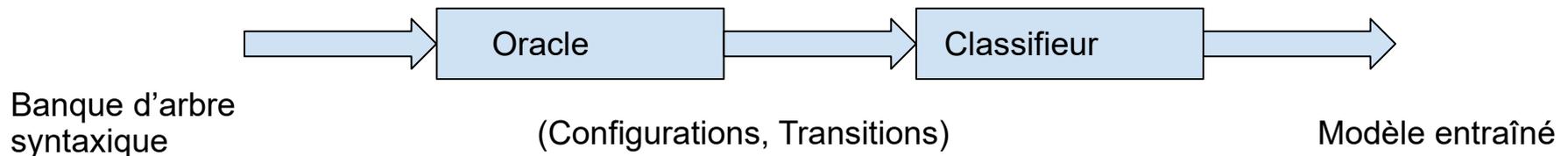
- Transition-Based Parser
- Automate = $\langle P, B, A \rangle$
- Transitions Arc Eager
 - Shift, Reduce, Right_label, Left_label

Construction de l'arbre syntaxique



Classifieur

- 1000 neurones
- 263 sorties
- Oracle = Automate TPB + Arbre syntaxique
- Entraînement :



Notre cas

- Le transfer parsing
 - Entraînement sur L
 - Test sur L0
 - Ressources limitées
- Caractéristiques typologiques
 - Description syntaxique des langues
 - Exploitation des similitudes syntaxiques
- Dé-lexicalisation

Ressources

- **Universal Dependencies (UD)**
 - Banque d'arbre syntaxique
 - 49 Langues
- **World Atlas of Language Structures (WALS)**
 - Description structurelles des langues
 - 2676 Langues
 - 192 caractéristiques

Wn et W80

- Wn
 - Naseem et al. (2012)
 - 6 caractères
 - 5 langues éliminées : 40 restantes
- W80
 - 80%
 - 22 caractéristiques pour 40 langues

Expériences

- L : Parseur standard
- Σ : Entraînement multi-lingue
- Σ +ID
- Σ +Wn
- Σ +W80

- Mesuré avec LAS

Résultats

- L vs Σ : -5,68
- L vs ΣID : -2,4
- Σ vs ΣWn : 2,77
- Σ vs $\Sigma W80$: 4
- $\Sigma W80$ vs ΣID : 0,72

L	Σ	ΣID	ΣWn	$\Sigma W80$
69,32	63,64	66,92	66,41	67,64

Travail personnel

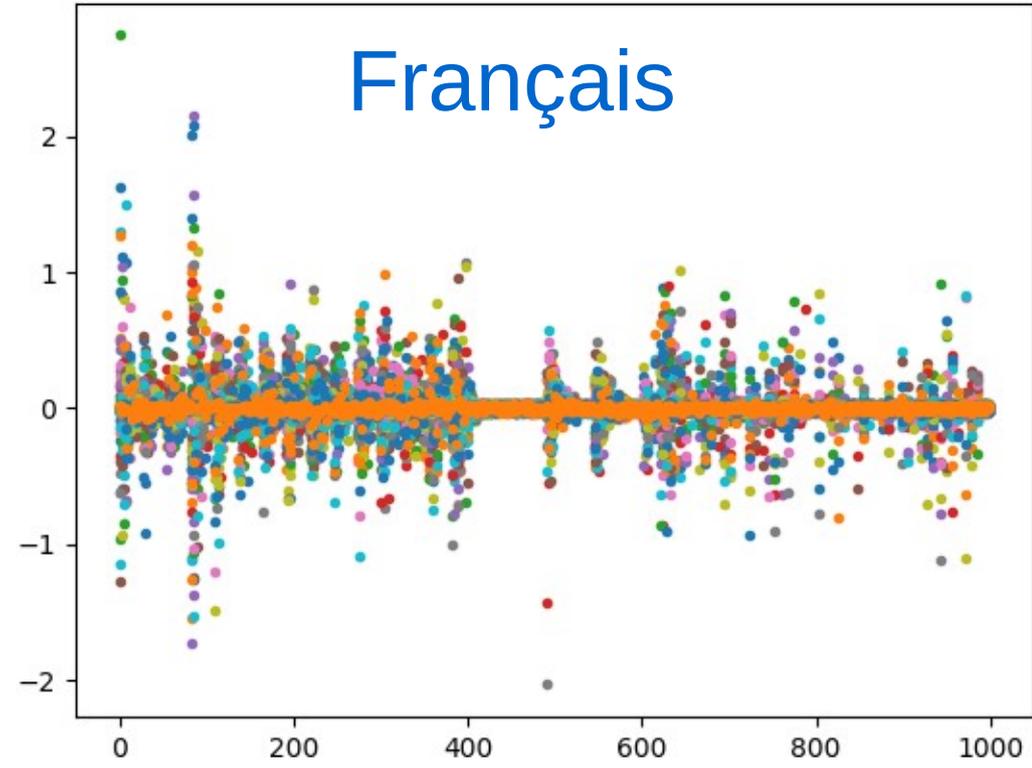
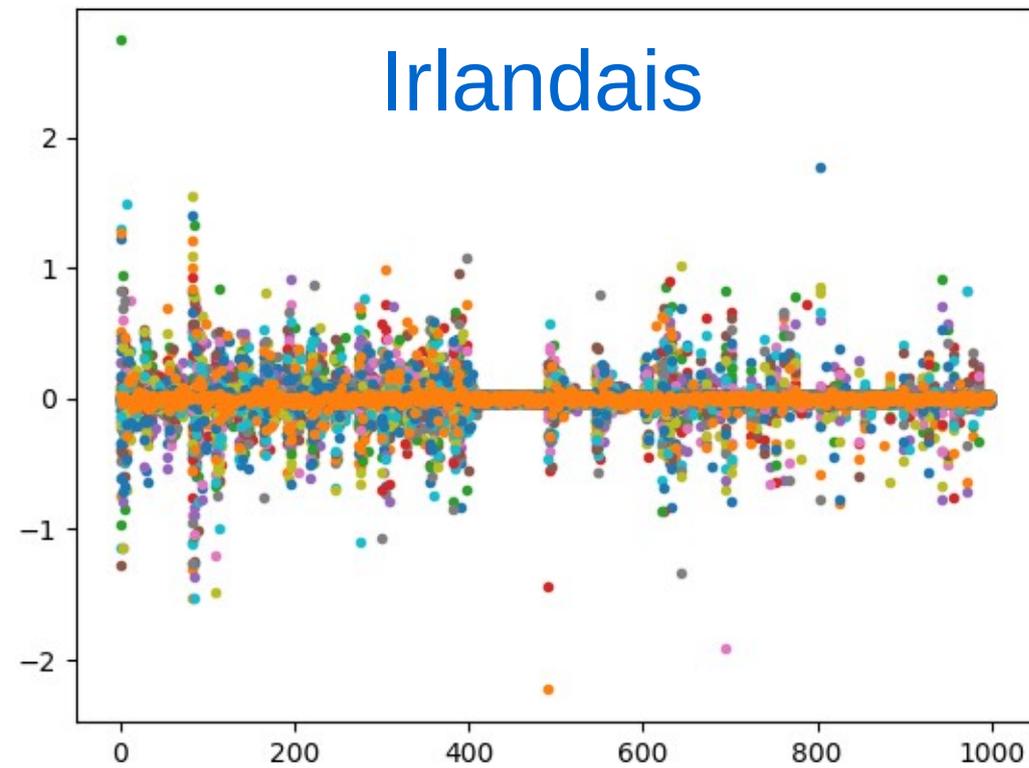
- Hypothèses :
 - Le réseau utilise W80 comme un identifiant de la langue.
 - Le réseau extrait de W80 des informations sur la façon dont la langue est construite.

Expérience

- Observer les poids des neurones pour différentes langues
- Neurone ID

Résultats

Poids entre W80 et les neurones



Sources

- Typological Features for Multilingual Delexicalised Dependency Parsing
- Wikipédia : Arbre syntaxique
- Tahira Naseem, Regina Barzilay, and Amir Globerson. 2012. Selective sharing for multilingual dependency parsing