

Détection de phénomènes syntaxiques à l'aide de modèles neuronaux

Alexis Nasr

8 novembre 2023

L'objectif de ce stage est de développer un modèle de détection de certains phénomènes syntaxiques dans des phrases.

Dans sa version la plus simple le modèle prendra en entrée une phrase et réalisera une classification binaire indiquant la présence ou l'absence d'un phénomène particulier. On aura donc autant de classifieurs binaires que de phénomènes à détecter.

Une seconde version réalisera de l'étiquetage multi-étiquettes et produira l'ensemble des phénomènes présents dans la phrase.

La troisième version produira en plus de chaque étiquette le début et la fin du segment contenant le phénomène d'intérêt.

On s'intéressera dans un premier temps à dix phénomènes décrits plus loin.

Pour apprendre de tels classifieurs, on utilisera des données annotées en syntaxe, plus particulièrement les données de Universal Dependencies [de Marneffe 2021]. Ces dernières associent à chaque phrase d'un corpus, l'analyse syntaxique de cette dernière sous la forme d'un arbre de dépendances. A l'aide du logiciel GREW [Guillaume 2021], on écrira des règles permettant de repérer dans la structure syntaxique d'une phrase la présence des phénomènes syntaxiques d'intérêt et d'associer ainsi aux phrases les étiquettes correspondant aux phénomènes présents dans la phrase. Ces données seront ensuite fournies à un classifieur qui apprendra à prédire les étiquettes correspondant aux phénomènes.

Différents modèles de prédiction seront testés, en particulier des réseaux récurrents et des réseaux fondés sur l'attention.

On s'intéressera dans un premier temps aux dix phénomènes suivants

1. SNL (Sujet Non Local) La tête du sujet est séparée du verbe par un certain nombre de mots.

Exemples :

- Le **livre** de Pierre **est** intéressant.
- Ce **livre**, très intéressant, **doit** être acheté.
- Ce **livre**, qui m'a été recommandé par un bon ami, **est** introuvable.

2. ONL (Objet Non Local) La tête de l'objet direct est séparée du verbe par un certain nombre de mots.

Exemples :

- **Je** propose à Pierre une **réunion** sur ce sujet.
- **Je** prends sans aucune hésitation le **plat** du jour.

3. SRC (Relative Sujet)

Exemples :

- J’ai reconnu le **garçon** qui **courait**.
 - J’ai reconnu la **fil**le, qui **courait**, sans difficulté.
4. ORC (Relative Objet)
- Exemples :
- J’ai reconnu le **garçon** que tu m’as présenté.
 - J’ai reconnu ce **garçon**, que tu m’as d’ailleurs déjà présenté
5. ORCL (LongORC) Contrôles à distance longs et non standard
- Exemples :
- J’ai aimé le **livre** que tu m’as dit que Jean t’avait demandé d’**acheter**.
 - Le **livre** que je suis allé à la FNAC lundi pour **acheter**
6. XRC locatives, dont, . . . :
- Exemples :
- La **ville** où il voulait **aller**.
 - La **conclusion** à laquelle il voulait en **venir**.
 - Le **livre** dont tu m’as dit qu’il fallait **parler**.
 - Le **livre** dont tu m’as dit qu’il **était** intéressant.
7. SCL clivée sujet
- Exemples :
- C’est **Paul** qui **mange**.
8. OCL clivée objet
- Exemples :
- C’est **Paul** que Marie **prévient**.
9. XCL clivée autre
- Exemples :
- C’est à **Paul** que Marie **offre** un cadeau.
10. CLI constructions avec double clitique complément
- Exemples :
- Il **lui en** parle.
 - Il **le lui** dit.

Bibliographie

- [Guillaume 2021] Bruno Guillaume. Graph Matching and Graph Rewriting : GREW tools for corpus exploration, maintenance and conversion. Demonstrations – 16th Conference of the European Chapter of the Association for Computational Linguistics (EACL).
- [de Marneffe 2021] Marie-Catherine de Marneffe, Christopher Manning, Joakim Nivre, Daniel Zeman (2021) : Universal Dependencies. In : Computational Linguistics, ISSN 1530-9312, vol. 47, no. 2, pp. 255-308.