

Titre du projet de programmation TER

## **Apprentissage automatique de grammaires à partir d'exemples positifs**

Tuteur : Christian Retoré <http://www.lirmm.fr/~retore>

L'apprentissage de grammaires formelles se présente dans de nombreux domaines : diagnostic, génomique, langage naturel,... Il s'agit à partir d'exemples donnés comme corrects d'inférer une grammaire qui produise ces exemples. Les exemples peuvent être des chaînes, des arbres, des graphes,... bref toutes les structures que des grammaires formelles peuvent produire.

Pour une sous classe des grammaires algébriques, les grammaires catégorielles rigides, Buszkowski et Penn ont proposé en 1990 un algorithme d'apprentissage de cette classe appelé RG à partir d'arbres aux branchements étiquetés. Cet algorithme utilise principalement l'unification des catégories grammaticales associées aux mots. Il a été montré que si les exemples fournis à l'algorithme RG sont effectivement l'énumération par une grammaire catégorielle rigide  $G$ , alors l'algorithme RG trouve une grammaire  $G'$  équivalente à  $G$  au bout d'un nombre fini d'exemples.

Le projet consiste à programmer cet algorithme RG et à le tester à partir d'exemples. On pourra aussi donner une grammaire  $G$ , engendrer les exemples de manière aléatoire et comparer la grammaire apprise  $G$  à la grammaire initiale  $G$ .

- Wojciech BUSZKOWSKI and Gerald PENN Categorical Grammars determined from linguistic data by unification. *Studia Logica* 49:431—454, 1990.
- Richard MOOT and Christian RETORÉ Classical Categorical Grammars: AB Grammars. In: *The Logic of Categorical Grammars. Lecture Notes in Computer Science*, vol 6850. Springer, 2012. [https://link.springer.com/chapter/10.1007/978-3-642-31555-8\\_1](https://link.springer.com/chapter/10.1007/978-3-642-31555-8_1)
- Jacques NICOLAS Grammatical inference as unification, RR-3632 INRIA, 1999. <https://hal.inria.fr/inria-00073042>