

Apprentissage de patrons lexico-syntaxiques à partir de textes

Valentina Dragos*, Marie-Christine Jaulent*

* INSERM, UMRS 872, eq. 20
15, Rue de l'école de médecine
75006 Paris
valentina.dragos@upmc.fr
marie-christine.jaulent@upmc.fr

Résumé. Ce papier présente une approche d'apprentissage de patrons lexico-syntaxiques à partir de textes annotés. Les patrons lexico-syntaxiques sont utilisés pour identifier des relations lexicales dans les corpus textuels. Leur construction manuelle est une tâche fastidieuse et des solutions permettant l'apprentissage sont souhaitables. Nous proposons une approche d'apprentissage qui repose sur l'utilisation des chemins de dépendance pour représenter les patrons et l'implémentation d'un algorithme de classification. L'approche a été appliquée dans le domaine biomédical pour identifier des patrons lexico-syntaxiques exprimant des relations fonctionnelles.

1 Introduction

Les patrons lexico-syntaxiques sont des structures représentant des schémas récurrents du langage. Ils sont utilisés dans le domaine du traitement automatique de la langue pour repérer des schémas langagiers dont ils sont l'abstraction. Les informations extraites sont des relations lexicales, voir par exemple Hearst (1992) ou Chagnoux et al. (2008) ou des relations spécifiques à un domaine particulier, cf. Koike et al. (2005).

La construction des patrons lexico-syntaxiques suppose un travail d'abstraction réalisé en amont, qui vise à définir une relation lexicale particulière et à identifier les différents contextes d'apparition de cette relation. Les éléments véhiculant la relation sont ainsi mis en évidence et synthétisés sous la forme d'un patron lexico-syntaxique. La construction manuelle des patrons présente deux inconvénients. Le premier est lié à leur portabilité. En changeant le domaine ou le type de langage d'un corpus de nouveaux patrons doivent être définis. Le deuxième inconvénient est lié à la mise en oeuvre d'applications à grande échelle. Dans ce cas, identifier un ensemble de patrons est loin d'être une tâche triviale. Ces deux inconvénients ont ouvert la voie à des travaux visant l'apprentissage de patrons lexico-syntaxiques. Nous proposons une approche d'apprentissage de patrons lexico-syntaxiques à partir de textes annotés. L'approche est générique et indépendante du domaine d'application. Nous l'avons utilisée dans le domaine biomédical pour identifier des patrons exprimant des relations fonctionnelles.