

Réseaux Lexicaux, Traitement des Langues, et Données Liées Ouvertes

Gilles Sérasset

Univ. Grenoble Alpes, LIG, GETALP, F-38000 Grenoble, France

CNRS, LIG, GETALP, F-38000 Grenoble, France

gilles.serasset@imag.fr

Résumé. Ces dernières décennies, notre regard sur les données lexicales informatisées a beaucoup évolué. D’abord annexe lexicale d’une grammaire ou d’une application, les dictionnaires d’application sont devenues bases lexicales dans lesquelles s’agrégeaient les données de différents modules. L’effort suivant s’est concentré dans la normalisation du format, avec notamment un mouvement massif vers le tout XML. Le travail de normalisation des structures des lexiques a suivi ensuite. Mais, alors que les normes restent structurellement proches des dictionnaires originaux (vus comme une collection d’entrées organisées de manière arborescentes), ont émergé des modèles de lexiques pensés comme des graphes.

Parallèlement, les travaux dans le domaine du Web Sémantique nous ont donné les moyens de représenter, manipuler et surtout partager nos ressources lexicales. En adoptant une représentation en RDF (Resource Description Framework), ainsi que l’approche des données liées ouverte (Linked Open Data), nous avons enfin les moyens de lier, fusionner, parcourir l’ensemble des ressources lexicales *comme s’il ne s’agissait que d’une seule ressource*.

Dans cette présentation, en m’appuyant sur les travaux réalisés dans le cadre des projets Papillon, LexALP et DBnary, j’essaierai de montrer en quoi, au delà de l’effet de mode actuel, l’utilisation du format des données liées ouvertes, est l’étape suivante naturelle dans notre étude du lexique.

Abstract. In recent decades, we have greatly changed the way we think about lexical data. First seen as a lexical annex of a grammar or an application, the application dictionaries became lexical databases which aggregates data from different modules. The next effort was concentrated in the standardization of format, with a massive movement towards XML. The work on standardization of lexical structure followed then. But while standards remain structurally close to the original dictionaries (seen as a collection of entries organized as trees), we now think the lexicon as graphs.

Meanwhile, the Semantic Web movement has given us the means to represent, manipulate, and share our lexical resources. By adopting RDF (Resource Description Framework) representation, and the Linked Open Data approach, we finally have the means to link, merge, browse all lexical resources *as if they were an unique resource*.

In this presentation, with the help of work done under the Papillon, LexALP and DBnary projects, I will try to show how, beyond the hype, Lexical Linked Open Data is the natural next step in our study of the lexicon.

Mots-clés : Données liées ouvertes, lexiques, traitement des langues.

Keywords: Linked Open Data, Lexicon, Natural Language Processing.