

**Un éditeur de définitions formelles pour les
connaissances lexicales de la théorie Sens-Texte**
Maxime Lefrançois, Fabien Gandon, Alain Giboin, Romain Gugert

► **To cite this version:**

Maxime Lefrançois, Fabien Gandon, Alain Giboin, Romain Gugert. Un éditeur de définitions formelles pour les connaissances lexicales de la théorie Sens-Texte. Catherine Faron-Zucker. IC - 25èmes Journées francophones d'Ingénierie des Connaissances, May 2014, Clermont-Ferrand, France. pp.279-281, 2014. <hal-00967537>

HAL Id: hal-00967537

<https://hal.inria.fr/hal-00967537>

Submitted on 28 Mar 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Un éditeur de définitions formelles pour les connaissances lexicales de la théorie Sens-Texte

Maxime Lefrançois^{1,2}, Fabien Gandon^{2,1}, Alain Giboin^{2,1}, Romain Gugert

¹ Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France

² Inria

{ maxime.lefrancois | fabien.gandon | alain.giboin }@inria.fr
romain.gugert@gmail.com

Résumé : Nous menons une étude en IC appliquée aux connaissances lexicales sémantiques de la Théorie Sens-Texte (TST). Avant de développer un formalisme adapté -le formalisme des Graphes d'Unités-, nous avons dû étendre la conceptualisation des prédicats linguistiques et des définitions lexicographiques dans la TST, puis nous avons voulu valider cette nouvelle conceptualisation auprès des lexicographes du projet RELIEF. Nous présentons donc un prototype d'éditeur qui permet de représenter formellement à l'aide de graphes des définitions lexicographiques. Initialement adaptées aux lexicographes, les Graphes d'Unités et une évolution du prototype d'éditeur pourraient être utilisés dans d'autres contextes de l'IC.

Mots-clés : Sémantique lexicale, définitions lexicographiques, Lexicographie, Théorie Sens-Texte, éditeur de graphes, formalisme des Graphes d'Unités.

1 Introduction

Nous menons une étude en IC appliquée aux connaissances lexicales sémantiques de la Théorie linguistique Sens-Texte (TST) (Mel'čuk *et al.*, 1995). Nous nous intéressons en particulier aux définitions lexicographiques, symbolisées par un réseau sémantique dans la TST.

Afin de faciliter le développement d'un formalisme de représentation des connaissances adapté -le formalisme des Graphes d'Unités (Lefrançois & Gandon, 2013)-, nous avons proposé une extension de la conceptualisation de la TST. Cette conceptualisation est intéressante pour d'autres applications en IC, car elle estompe la distinction habituelle concept/relation.

Nous présentons un prototype d'éditeur qui permet de représenter formellement à l'aide de graphes des définitions lexicographiques selon cette conceptualisation étendue, que nous avons fait évaluer par des lexicographes¹ du projet RELIEF² (Lux-Pogodalla & Polguère, 2011).

Cet article présente la conceptualisation actuelle des définitions lexicographiques dans le projet RELIEF (§2), l'extension de cette conceptualisation que l'on propose (§3), puis le prototype d'éditeur développé (§4). Nous concluons par son évaluation et les perspectives (§5).

2 Les définitions lexicographiques dans la théorie Sens-Texte

La définition lexicographique d'une unité lexicale L présente de façon formelle le sens dénotationnel de L . Par exemple PEIGNE_{2A} (le peigne du tisserand) peut être défini par :

$peigne_{2a}$ de X pour $Y = (\text{Outil de tissage qu'une personne } X \text{ utilise pour démêler les fibres d'un objet } Y)$

1. La lexicographie est une science qui a pour sujet d'étude l'édition des dictionnaires

2. RELIEF est un projet d'envergure de l'ATILF, CNRS - <http://www.atilf.fr/>

Dans le projet RELIEF, l'édition d'une définition lexicographique s'effectue en trois étapes (c.f., fig. 1, gauche). Cependant, Wanner (2003) note qu'il est souhaitable de formaliser d'avantage les définitions lexicographiques, en particulier pour des applications de TALN. Une formalisation en vue est sous la forme de réseaux sémantiques (Mel'čuk *et al.*, 1995).

3 Extension de la conceptualisation

Nous avons introduit un niveau sémantique profond pour conceptualiser les sens. On peut y organiser les prédicats sémantiques en une hiérarchie au sein de laquelle des positions actanciennes obligatoires, optionnelles ou interdites, et munies de signatures, sont héritées et potentiellement spécialisées (Lefrançois & Gandon, 2013).

Les définitions lexicographiques sont conceptualisées à ce niveau sous forme de multigraphes étiquetés et orientés, dont la visualisation s'inspire de l'UML.

4 Prototypage de définitions lexicographiques formelles

Nous avons proposé un workflow en quatre étapes adapté à la nouvelle conceptualisation (Lefrançois *et al.*, 2013) (c.f., fig. 1, droite). L'éditeur dont nous présenterons une démonstration³ est une implémentation de ce workflow (Gugert, 2013).

Processus suivi dans le projet RELIEF 1. Sélection d'une étiquette sémantique dans une hiérarchie (Polguère, 2011) 2. Sélection de la structure actancielle sémantique de surface (Mel'čuk, 2004) 3. Rédaction de la définition lexicographique en XML (Barque <i>et al.</i> , 2010)	Processus à suivre avec le nouvel éditeur 1. Sélection de la structure actancielle de la sémantique profonde 2. Édition de la définition formelle 3. Sélection de la structure actancielle sémantique de surface 4. Mise en correspondance des structures actanciennes
--	---

TABLE 1 – à g., workflow des lexicographes du projet RELIEF ; à dr., workflow de l'éditeur.

Pour l'utilisateur, les opérations d'édition de la définition (c.f., fig. 1b) sont des opérations de manipulation de graphes. Le "glisser-déposer" lui permet de sortir une position actancielle, pour éventuellement se rendre compte que sa signature comporte elle-même des positions actanciennes, ou pour fusionner des nœuds. Nous nous sommes inspirés de l'UML, mais nous ne connaissons pas d'éditeur d'UML qui permette de "sortir" un attribut à l'extérieur de sa classe pour ainsi obtenir une association vers une autre classe.

Techniquement, ces opérations sont implémentés en JavaScript au dessus de *mxGraph*⁴, qui permet la visualisation et la manipulation de graphes.

5 Évaluation et perspectives

Nous avons évalué l'acceptation du prototype et du processus supportant la nouvelle conceptualisation par la communauté de la TST. Les résultats sont encourageants (Gugert, 2013) et per-

3. Démonstration de l'éditeur - <http://wimmics.inria.fr/doc/video/UnitGraphs/editor1.html>

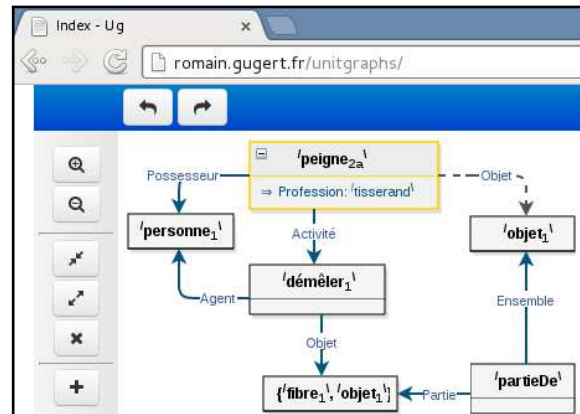
4. *mxGraph* - visualisation de graphes en JavaScript - <http://www.jgraph.com/mxgraph.html>

mettent d'entrevoir des directions d'amélioration : nous devons en particulier rendre le workflow plus semblable à celui qui est utilisé actuellement.

Initialement adaptées aux lexicographes, les étapes 1 et 2 peuvent être réutilisées dans un autre contexte d'IC. Par ailleurs, une évolution de ce prototype devrait guider l'utilisateur dans l'élaboration de la structure actancielle en même temps qu'il manipule le graphe de définition.



(a) Phase 1 : Sélection de la structure actancielle.



(b) Phase 2 : Édition de la définition formelle par manipulation de graphes.

FIGURE 1 – Captures d'écran du prototype lors de l'édition de la définition de PEIGNE_{2A}.

Références

- BARQUE L., NASR A. & POLGUÈRE A. (2010). From the Definitions of the 'Trésor de la Langue Française' To a Semantic Database of the French Language. In FRYSKA AKADEMY, Ed., *Proceedings of the XIV Euralex International Congress*, Fryske Akademy, p. 245–252, Leeuwarden, Pays-Bas.
- GUGERT R. (2013). Scénarisation d'interactions avec les objets du formalisme des Graphes d'Unités et prototypage d'un éditeur de définitions lexicographiques formelles. Mémoire de Master 2 - UPMC, hal-00860767.
- LEFRANÇOIS M. & GANDON F. (2013). Rationale, Concepts, and Current Outcome of the Unit Graphs Framework. In *Proceedings of the 9th International Conference on Recent Advances in Natural Language Processing (RANLP 2013)*, p. 382–388.
- LEFRANÇOIS M., GUGERT R., GANDON F. & GIBOIN A. (2013). Application of the Unit Graphs Framework to Lexicographic Definitions in the RELIEF project. In *Proceedings of the 6th International Conference on Meaning-Text Theory (MTT'2013)*, Prague, Czech Republic.
- LUX-POGODALLA V. & POLGUÈRE A. (2011). Construction of a French Lexical Network : Methodological Issues. In *Proceedings of the International Workshop on Lexical Resources*, Ljubljana.
- MEL'ČUK I. (2004). Actants in semantics and syntax I : Actants in semantics. *Linguistics*, **42**(1), 247–291.
- MEL'ČUK I., CLAS A. & POLGUÈRE A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Paris/Louvain-la-Neuve : Duculot.
- POLGUÈRE A. (2011). Classification sémantique des lexies fondée sur le paraphrasage. *Cahiers de lexicologie*, **98**, 197–211.
- WANNER L. (2003). Definitions of Lexical Meanings : Some Reflections on Purpose and Structure. In *Proceedings of the First international conference on Meaning-Text Theory (MTT'2003)*, p. 16–28.