

Méthodologie d'ingénierie des connaissances pour la représentation des définitions lexicographiques dans le cadre de la théorie Sens-Texte

Maxime Lefrançois, Fabien Gandon, Alain Giboin

► To cite this version:

Maxime Lefrançois, Fabien Gandon, Alain Giboin. Méthodologie d'ingénierie des connaissances pour la représentation des définitions lexicographiques dans le cadre de la théorie Sens-Texte. TOTH - 8th International Conference on Terminology

Ontology: Theories and applications, Jun 2014, Chambéry, France. Institut Porphyre, Savoir et Connaissance, 2014. <hal-01061399>

HAL Id: hal-01061399

<https://hal.inria.fr/hal-01061399>

Submitted on 5 Sep 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.

Méthodologie d'ingénierie des connaissances pour la représentation des définitions lexicographiques dans le cadre de la théorie Sens-Texte

Maxime Lefrançois*, Fabien Gandon**, Alain Giboin**

*Univ. Nice Sophia Antipolis, CNRS, I3S, UMR 7271, 06900 Sophia Antipolis, France
maxime.lefrancois.86@gmail.com

**Inria

[fabien.gandon, alain.giboin]@inria.fr

Résumé. La Théorie Sens-Texte (TST) est l'une des théories linguistiques les plus précises d'un point de vue descriptif. En représenter les connaissances est un défi pour des applications de lexicographie, mais également pour injecter des connaissances linguistiques fines dans les systèmes de TALN. Cet article décrit la méthodologie et les principaux résultats de notre recherche qui consiste à mener une étude d'Ingénierie des Connaissances pour modéliser les connaissances lexicales sémantiques de la TST et en formaliser la sémantique. Nous nous intéressons plus particulièrement aux prédicats linguistiques, aux représentations linguistiques, et aux définitions lexicographiques. Nous nous sommes inspirés de la méthodologie de Bachimont, et avons opéré en trois étapes principales : (i) extension de la conceptualisation du domaine, (ii) recherche d'un formalisme de représentation des connaissances adapté, (iii) opérationnalisation du formalisme pour le raisonnement logique et pour le partage et l'interrogation des connaissances sur le web des données.

1 Introduction

En tant que théorie linguistique hautement précise du point de vue descriptif, la Théorie Sens-Texte (TST) tient une place centrale dans le champ de la sémantique lexicale (Mel'čuk *et al.*, 1995; Kahane, 2003). La richesse de description des fonctions lexicales et des définitions lexicographiques du Dictionnaire Explicatif et Combinatoire (DEC) fait sa force, mais elle a aussi des limites : (i) l'absence de représentation formelle et d'axiomatisation ; (ii) la faible couverture de la langue ; et (iii) le peu d'applications en TALN.

Lorsque Mel'čuk *et al.* (1995, p.208) promeut l'informatisation du DEC dans le but de “*faire tourner*” un *appareillage logico-déductif sur cette description*”. Il met notamment en évidence l'émergence de besoins spécifiques :

- vérifier la validité du lexique ;
- découvrir de nouvelles connaissances sur la langue ;
- raisonner sur les textes.

L'émergence de tels besoins existe en réalité lors de la production de connaissances par tout domaine de spécialité. L'Ingénierie des Connaissances (IC) vise justement à répondre à ces besoins. Il s'agit de permettre par exemple de représenter, de manipuler, d'échanger, d'interroger, ou de raisonner avec les connaissances. Les besoins envisagés par Mel'čuk seraient donc couverts.

En parallèle aux travaux de formalisation déjà menés sur le DEC, nous nous sommes donc proposés de représenter les prédicats linguistiques, les représentations linguistiques, et les définitions lexicographiques au sens de l'IC, en restant au plus proche des standards de la Représentation des Connaissances (RC).

Dans cet article nous rapportons la méthodologie et les principaux résultats de notre recherche en ingénierie des connaissances appliquée aux connaissances lexicales sémantiques du DEC. La section 2 introduit la notion de définition lexicographique, à laquelle nous nous intéressons plus particulièrement. Nous détaillons ensuite en section 3 notre méthodologie, qui s'inspire de celle proposée par Bachimont (2000).

2 Les Définitions Lexicographiques dans le cadre de la TST

La définition lexicographique d'une unité lexicale L présente de façon formelle le sens dénotatif de L. Dans le DEC, les définitions lexicographiques sont analytiques, c'est à dire de type genre prochain et différences spécifiques. Par exemple, nous pouvons trouver à l'entrée PEIGNE_{B.2.D} (le peigne du tisserand) la définition :

peigne_{B.2.d} de personne X pour objet Y =
(Outil de tissage qu'une personne X utilise pour démêler₂ les fibres d'un objet Y)

Le membre gauche de l'égalité représente le défini, PEIGNE_{B.2.D}, sous sa forme propositionnelle : une expression élémentaire dans laquelle apparaissent PEIGNE_{B.2.D} et sa *structure actancielle sémantique*, définie selon la théorie des Actants Sémantiques (Mel'čuk, 2004). Le membre droit de l'égalité représente le définissant. Le genre prochain de (peigne_{B.2.d}) est (outil), et le reste de la définition précise les différences spécifiques de (peigne_{B.2.d}) par rapport à (outil). La structuration des composantes sémantiques de la définition est encore discutée par les chercheurs en lexicologie explicative et combinatoire aujourd'hui (Iordanskaja et Mel'čuk, 1990; Mel'čuk *et al.*, 1995; Mel'čuk et Wanner, 2001; Altman et Polguère, 2003; Barque, 2008; Barque et Polguère, 2009).

Cependant, il est nécessaire d'aboutir à une structuration aussi formelle que possible des définitions lexicographiques pour les applications en TALN, comme le justifie Wanner (2003). C'est pourquoi nous nous intéressons à la conceptualisation originellement privilégiée dans la TST, qui présente les définitions lexicographiques sous la forme d'une représentation sémantique qui encode le sens de l'unité lexicale définie (Mel'čuk *et al.*, 1995). Par exemple la figure 1 représente la définition de PEIGNE_{B.2.D}. Le nœud étiqueté (outil) est marqué comme communicativement dominant, il s'agit du genre prochain de PEIGNE_{B.2.D}. D'autres nœuds sont marqués X, Y,... et représentent la structure actancielle sémantique de PEIGNE_{B.2.D}.

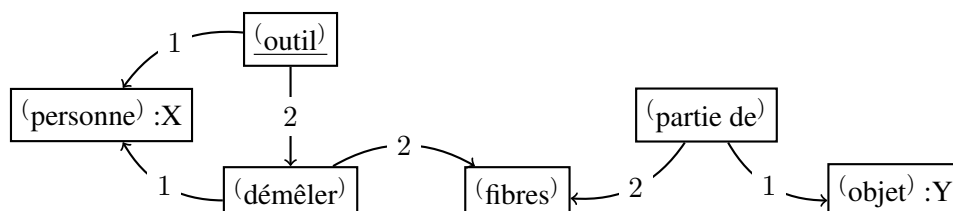


FIG. 1 – Représentation de la définition lexicographique de PEIGNE_{B.2.D} sous la forme d'une représentation sémantique

Jusqu'à présent, et de par la difficulté pratique de rédiger -à la main- un ensemble cohérent de telles représentations (Mel'čuk *et al.*, 1995, p.75), ce type de représentation n'est pas utilisé dans les projets d'informatisation du DEC (Altman et Polguère, 2003; Apresian *et al.*, 2003; Wanner, 2009; Lux-Pogodalla et Polguère, 2011). Néanmoins, lorsque Polguère (1990) précise le système de règles dans la composante sémantique du modèle Sens-Texte dans une perspective de génération automatique de textes, il justifie l'importance de représenter les définitions lexicographiques sous cette forme. On peut en effet en déduire deux règles de paraphrase qui devraient permettre de transformer une représentation sémantique en une autre équivalent, en remplaçant une unité sémantique par sa paraphrase, et vice-versa. Ces opérations élémentaires au niveau sémantique sont fondamentales pour faciliter le choix d'une lexicalisation appropriée pour un ensemble de représentations sémantiques équivalents.

3 La méthodologie

Lorsque Mel'čuk *et al.* (1995, p.208) promeut l'informatisation du DEC, il met en évidence l'émergence de besoins spécifiques, tels que la recherche intelligente d'information, la mise à jour et la vérification automatique, ou encore la déduction logique. L'émergence de tels besoins existe en réalité lors de la production de connaissances par tout domaine de spécialité. L'ingénierie des connaissances se propose justement de répondre à ces besoins. Il s'agit de permettre par exemple de représenter, de manipuler, d'échanger, d'interroger, ou de raisonner avec les connaissances.

La tâche de l'ingénierie des connaissances, telle que Bachimont (2000) la définit, est de modéliser formellement un problème pour lequel les seules connaissances dont on dispose sont de nature linguistiques ou cognitives. La sortie de ce processus de modélisation est un formalisme de représentation de connaissances, qui peut alors être mis en œuvre par les machines. Nous nous sommes donc inspirés de la méthodologie en trois étapes proposée par Bachimont (2000), pour passer de l'expression linguistique des connaissances vers une représentation formelle et calculable.

3.1 Extension de la conceptualisation

La première étape consiste, à partir des connaissances de TST, à définir les primitives du domaine. Bien que la théorie des actants sémantiques et les définitions lexicographiques soient déjà hautement conceptualisés, nous avons identifié certaines limitations pour une formalisation ultérieure.

Conceptualisation des prédicats linguistiques. Pour une unité lexicale L, Mel'čuk (2004) distingue L dans la langue (i.e., dans le dictionnaire), ou dans les textes (i.e., dans les énoncés). Pour désactiver cette ambiguïté, nous nommons *type d'unité lexicale* l'unité lexicale décrite dans le dictionnaire, et *instance d'unité lexicale* (ou simplement *unité lexicale*, l'unité lexicale décrite dans les textes. De plus, nous associons à chaque type d'unité sémantique *de surface* une structure actancielle qui correspond à la structure actancielle sémantique de son type de lexie associé.

Nous avons montré (Lefrançois et Gandon, 2013c) que la hiérarchisation des prédicats sémantiques, avec héritage et spécialisation de leur structure actancielle, ne peut correspondre à une hiérarchisation des sens. Notons qu'en nous plaçant dans une démarche d'IC, nous ne souhaitons en aucun cas modifier la conceptualisation existante, mais seulement l'étendre. Nous avons donc introduit un nouveau niveau dit *sémantique profond* pour représenter les sens, et y avons défini les notions d'unité sémantique profonde et de leur structure actancielle de sorte qu'ils puissent être organisés en une hiérarchie qui satisfait les propriétés suivantes :

1. la hiérarchie des types d'unité sémantique profonde représente une hiérarchisation des sens ;
2. la structure actancielle des types d'unité sémantique profonde est héritée et éventuellement spécialisée au sein de la hiérarchie ;
3. il existe une correspondance simple entre la structure actancielle des types d'unité sémantique profonde, et celle des types d'unité sémantique de surface.

Finalement, nous proposons une visualisation graphique inspirée d'UML, comme l'illustre la figure 2. Les rectangles du haut représentent un type de lexie, un type d'unité sémantique de surface doté de sa structure actancielle, et un type d'unité sémantique profonde doté de sa structure actancielle. Les rectangles du bas représentent des instances de ces types.

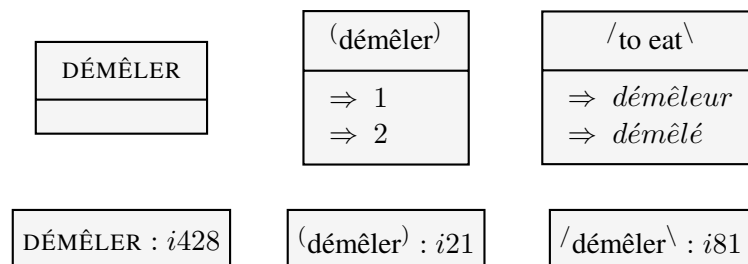


FIG. 2 – Visualisation des types et instances d'unité linguistiques.

Conceptualisation des définitions lexicographiques. Une représentation d'une définition en langue naturelle, aussi formalisée soit-elle, reste une linéarisation qui encode le sens de la lexie définie. D'un autre côté, une définition lexicographique représentée sous la forme d'une représentation sémantique met en scène des (instances d')unités sémantiques. Il s'agit donc d'une instanciation prototypique de la définition d'un lexie L. Nous avons donc proposé une nouvelle conceptualisation pour les définitions lexicographiques :

- au niveau sémantique profond,
- et au niveau du dictionnaire.

La figure 3 représente la reconceptualisation de la définition de la lexie PEIGNE_{B.2.D}.

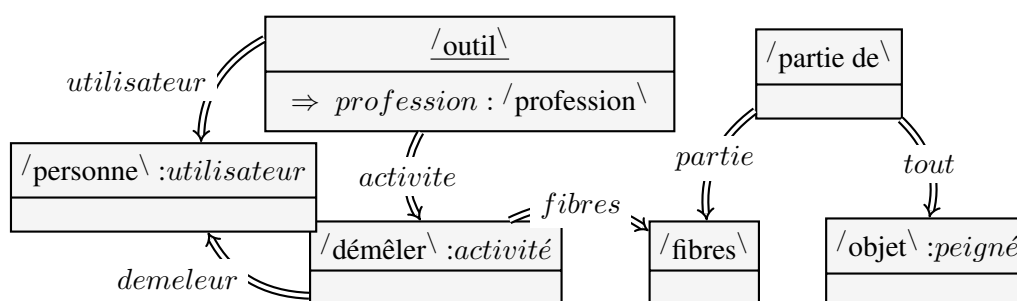


FIG. 3 – Repositionnement de la définition lexicographique de PEIGNE_{B.2.D} au niveau sémantique profond et au niveau du dictionnaire.

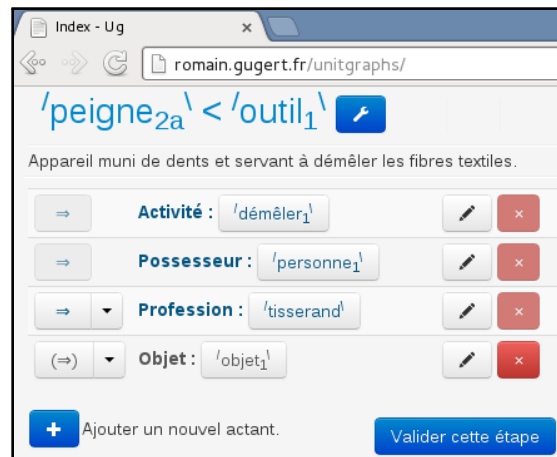
3.2 Développement d'un prototype d'éditeur de définitions lexicographiques pour assister la conceptualisation

Nous avons proposé un workflow d'édition d'une définition lexicographique selon cette conceptualisation étendue (Lefrançois *et al.*, 2013), et développé un prototype d'éditeur qui implémente ce workflow (Gugert, 2013; Lefrançois *et al.*, 2014). Une démonstration est disponible à l'adresse <http://wimmics.inria.fr/doc/video/UnitGraphs/editor1.html>. La figure 4 représente les deux premières étapes du processus d'édition d'une définition lexicographique dans ce prototype :

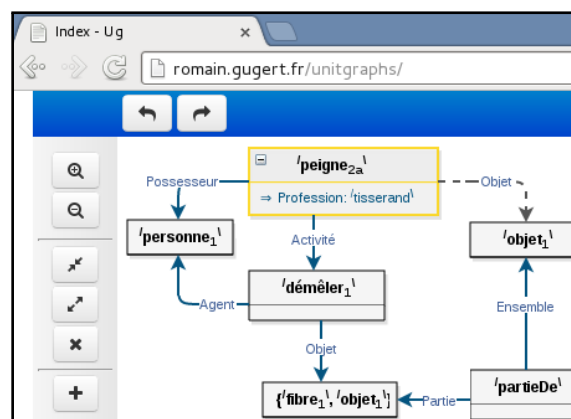
1. choix du genre prochain et de la structure actancielle sémantique profonde ;
2. construction de la définition lexicographique formelle par manipulation directe de graphes.

Nous avons évalué l'acceptation du processus et du prototype par la communauté de la TST. Nous avons ainsi pu confirmer les attentes des lexicographes de la TST pour un tel outil pratique, et identifier différentes directions d'amélioration pour notre prototype. Finalement, ce prototype offre une première base essentielle sur laquelle de

futures discussions peuvent s'appuyer, et il permettra, nous l'espérons, l'intégration future de nos travaux dans des logiciels de lexicographie.



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



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

FIG. 4 – Captures d'écran du prototype lors de l'édition de la définition de PEIGNE_{B.2.D}.

3.3 Elaboration d'un formalisme de représentation des connaissances adapté

L'extension de la conceptualisation ainsi proposée peut être vue comme un objet frontière de type forme standardisée, au sens de Star et Griesemer (1989), entre la TST et le domaine de la représentation des connaissances.

Une fois la conceptualisation de la TST étendue, nous sommes passés à l'étape de formalisation des connaissances. Il s'agit de concevoir une ontologie du domaine, sa conception étant intimement liée au formalisme de représentation des connaissances que l'on choisit. Afin d'optimiser les possibilités de raisonnement dans le formalisme, notre étude s'est limitée à l'utilisation de fragments décidables de la logique du premier ordre. Ainsi, au premier abord, deux formalismes de représentation des connaissances existants semblaient adaptés pour la représentation des définitions lexicographiques :

- les logiques de description, qui sont intégrées aux standards de facto de la représentation des connaissances : les formalismes du Web Sémantique ;
- le formalisme des Graphes Conceptuels, qui avait originellement pour but la représentation des connaissances linguistiques à des fins de traitement automatisé, et qui fut inspiré des travaux de Tesnière (1959) à l'instar de la TST.

Montrons que ces deux formalismes ne sont pas adaptés pour représenter les connaissances de la TST à la lumière de deux des cinq critères de conception d'une ontologie proposés par Gruber (1995) : la clarté et la cohérence.

La clarté La définition d'un concept doit faire passer le sens voulu du terme, de manière aussi objective que possible. Dans la lignée de (Lefrançois et Gandon, 2011), nous avons montré que certains graphes simples peuvent être projetés sur la définition d'une seule classe en logique de description. Cependant, le manque de clarté de l'ensemble abscons d'axiomes ainsi obtenu justifie l'abandon de cette direction de recherche pour notre étude. Nous lui avons préféré dans un premier temps le formalisme des Graphes Conceptuels, qui propose une représentation des définitions de type de concepts et de relations très similaire aux définitions lexicographiques (Leclère, 1998).

La cohérence Rien qui ne puisse être inféré de l'ontologie ne doit entrer en contradiction avec les définitions des concepts. Bien qu'on puisse à priori représenter une définition lexicographique par un couple ⟨définition de type de concept, définition de type de relation⟩ dans le formalisme des GC, nous pouvons montrer que certaines inférences seraient incohérentes.

Nous avons donc décidé de modifier les bases du formalisme des Graphes Conceptuels, afin d'en dériver un nouveau formalisme, nommé formalisme des Graphes d'Unités (Lefrançois et Gandon, 2013a). La base de ce formalisme est une hiérarchie des types d'unité notée \mathcal{T} : l'ensemble minimal d'objets mathématiques nécessaires à la formation d'un ensemble préordonné de types d'unités munis de structures actanciennes obligatoires, optionnelles, ou interdites. Un Graphe d'Unité est alors une combinaison de nœuds d'unité interconnectés.

3.4 Opérationnalisation de la formalisation sur le Web des données

La dernière étape de la méthodologie de l'ingénierie des connaissances est d'opérationnaliser le formalisme. Il s'agit de déterminer dans quelle mesure le formalisme choisi peut être utilisé pour les scénarios d'utilisation effectivement envisagés.

Nous avons tout d'abord étudié la déduction dans le formalisme des Graphes d'Unités. Nous lui avons attribué une sémantique logique (Lefrançois et Gandon, 2013b,d), et avons étudié les conditions de décidabilité du problème de conséquence logique. Notons que ce problème est décidable sous des conditions d'acyclicité que les lexicographes de la TST ont l'habitude de s'imposer dans l'élaboration du dictionnaire (Lefrançois, 2014).

C'est seulement après ce travail que nous avons choisi une modélisation du formalisme des Graphes d'Unités à l'aide des formalismes du Web Sémantique, afin de profiter des architectures existantes pour le partage, l'interopérationnalisation, et l'interrogation des connaissances sur le Web des données (Lefrançois, 2014). Nous utilisons le métamodèle OWL 2 RL, et nous étendons la base de règles OWL 2 RL/RDF par un ensemble de règles SPARQL pour capturer la sémantique de notre formalisme. Nous utilisons deux modèles hétérogènes :

- le modèle **gu-langue** adopte le point de vue de la compétence, et décrit les types d'unité au niveau assertionnel. Il s'agit du point de vue adopté par les ontologies lexicales du Web des données.
- Le modèle **gu-usage**, au contraire, adopte le point de vue de la performance. Les types d'unité sont décrits au niveau terminologique, laissant le niveau assertionnel aux graphes d'unités.

La figure 5 illustre ces deux modèles. Notons que ces modèles hétérogènes sont rendus interopérables par un ensemble de règles de transfert, qui définissent un morphisme d'ontologies au sens de Flouris *et al.* (2008).

Méthodologie d'IC pour représenter les définitions lexicographiques de la TST

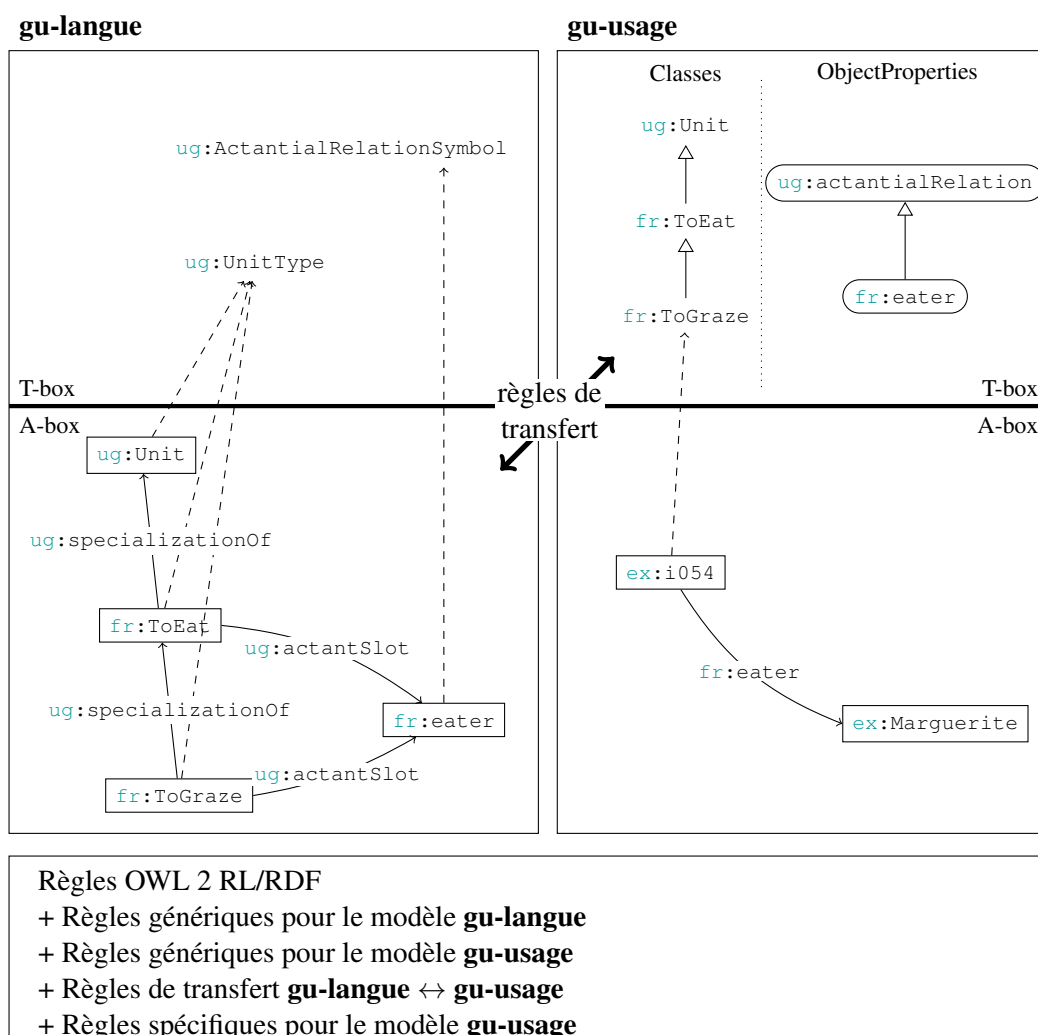


FIG. 5 – Architecture globale du modèle.

La sémantique du formalisme des Graphes d'Unités est alors capturée par un ensemble de règles SPARQL :

- les règles SPARQL OWL 2 RL/RDF écrites en SPARQL, accessibles à l'URL <http://ns.inria.org/ug/v1/owl2rl.rul> ;
- des règles complémentaires génériques pour le modèle **gu-langue**, accessibles à l'URL <http://ns.inria.org/ug/v1/ug-language.rul> ;
- des règles complémentaires génériques pour le modèle **gu-usage**, accessibles à l'URL <http://ns.inria.org/ug/v1/gu-usage.rul> ;
- des règles de transfert entre les deux modèles, accessibles à l'URL <http://ns.inria.org/ug/v1/ug-transfer.rul> ;
- les règles spécifiques d'expansion et de contraction des types d'unité formellement définis. Notons que nous avons proposé un modèle des définitions formelles, et des transformations SPARQL TEMPLATE pour générer des règles SPARQL d'expansion et de contraction. Par exemple la représentation de la définition lexicographique de PEIGNE_{B.2.D} est illustrée sur la figure 6, et la sortie du processus de transformation est illustrée sur la figure 7.

```

1 ex:peigneB2d ug:definition _:outil . 22      a ug:Participant ;
2                                          23      a ex:Fibres .
3 _:outil {                               24
4                                          25      _:partieDe rdfs:label "partieDe" ;
5      _:outil rdfs:label "outil" ;       26      a ug:Participant ;
6      a ug:Participant ;                 27      a ex:PartieDe .
7      a ex:Outil .                       28
8                                          29      _:objet rdfs:label "objet" ;
9      _:personne rdfs:label "personne" ;30      a ug:Participant ;
10     a ug:Participant ;                 31      a ex:Objet ;
11     a ex:Personne ;                   32      a ug:obligatoryParticipant ;
12     a ug:obligatoryParticipant ;     33      ug:marker ex:peigne .
13     ug:marker ex:utilisateur .        34
14                                       35      _:outil ex:utilisateur _:personne
15      _:demeler rdfs:label "demeler" ;  36      .
16     a ug:Participant ;                 37      _:outil ex:activite _:demeler .
17     a ex:Demeler ;                     38      _:demeler ex:demeleur _:personne .
18     a ug:obligatoryParticipant ;     39      _:demeler ex:fibres _:fibres .
19     ug:marker ex:activite .           40      _:partieDe ex:partie _:fibres .
20                                       41      _:partieDe ex:tout _:objet .
21      _:fibres rdfs:label "fibres" ;

```

FIG. 6 – Représentation RDF de la définition lexicographique de PEIGNE_{B.2.D}.

Méthodologie d'IC pour représenter les définitions lexicographiques de la TST

Règle d'expansion **def+**

```
1 # expansion rule
2 CONSTRUCT {
3   ?u a ex:Outil .
4   ?personne a ex:Personne .
5   ?activite a ex:Demeler .
6   ?fibres a ex:Fibres .
7   ?partieDe a ex:PartieDe .
8   ?peigne ex:Objet .
9   ?u ex:utilisateur ?personne .
10  ?u ex:activite ?activite .
11  ?activite ex:demeleur ?personne .
12  ?activite ex:fibres ?fibres .
13  ?partieDe ex:partie ?fibres .
14  ?partieDe ex:tout ?peigne .
15 }
16 WHERE {
17   ?u a ex:peigneB2d .
18   ?u ex:utilisateur ?personne .
19   ?u ex:activite ?activite .
20   ?u ex:peigne ?peigne .
21   FILTER NOT EXISTS {
22     ?u a ex:Outil .
23     ?personne a ex:Personne .
24     ?activite a ex:Demeler .
25     ?fibres a ex:Fibres .
26     ?partieDe a ex:PartieDe .
27     ?peigne ex:Objet .
28     ?u ex:utilisateur ?personne .
29     ?u ex:activite ?activite .
30     ?activite ex:demeleur ?personne .
31     ?activite ex:fibres ?fibres .
32     ?partieDe ex:partie ?fibres .
33     ?partieDe ex:tout ?peigne .
34   }
35 }
```

Règle de contraction **def-**

```
1 # contraction rule
2 CONSTRUCT {
3   ?u a ex:peigneB2d .
4   ?u ex:utilisateur ?personne .
5   ?u ex:activite ?activite .
6   ?u ex:peigne ?peigne .
7 }
8 WHERE {
9   ?u a ex:Outil .
10  ?personne a ex:Personne .
11  ?activite a ex:Demeler .
12  ?fibres a ex:Fibres .
13  ?partieDe a ex:PartieDe .
14  ?peigne ex:Objet .
15  ?u ex:utilisateur ?personne .
16  ?u ex:activite ?activite .
17  ?activite ex:demeleur ?personne .
18  ?activite ex:fibres ?fibres .
19  ?partieDe ex:partie ?fibres .
20  ?partieDe ex:tout ?peigne .
21  FILTER NOT EXISTS {
22    ?u a ex:peigneB2d .
23    ?u ex:utilisateur ?personne .
24    ?u ex:activite ?activite .
25    ?u ex:peigne ?peigne .
26  }
27 }
```

FIG. 7 – Règles d'expansion et de contraction générées automatiquement à partir de la représentation RDF de la définition lexicographique de PEIGNE_{B,2,D}.

Nous avons proposé une spécialisation du modèle des Graphes d'Unités pour la TST, que nous avons alignée avec le modèle **ontolex** proposé par le groupe communautaire Ontology-Lexica du W3C¹. Nous espérons retirer de cette approche :

- une adoption facilitée de nos travaux par la communauté de la représentation des connaissances lexicales ;
- la possibilité de réutiliser des données existantes pour peupler notre formalisme.

Nos résultats sont publiés en ligne à l'adresse <http://ns.inria.fr/ug/v1#>.

4 Conclusion

Ainsi, afin de rendre possible la représentation des définitions lexicographiques sous la forme d'une représentation sémantique dans un projet d'informatisation du Dictionnaire Explicatif et Combinatoire, nous avons appliqué une méthodologie en trois étapes :

- Nous avons proposé une extension de la conceptualisation de la théorie Sens-Texte, qui joue le rôle d'un objet frontière avec le domaine de l'ingénierie des connaissances ;
- Nous avons choisi un formalisme de représentation des connaissances adapté. Nous avons montré que les logiques de description et le formalisme des Graphes Conceptuels ne sont pas adaptés, et nous avons donc construit le nouveau formalisme de représentation des connaissances des Graphes d'Unités ;
- Nous avons étudié dans quelle mesure ce formalisme peut être opérationnalisé pour répondre aux besoins des linguistes. D'abord en étudiant la déduction logique et les conditions de décidabilité du formalisme, puis en proposant une modélisation avec les formalismes du Web Sémantique, afin de profiter des architectures existantes pour le partage, l'interopéationnalisé, et l'interrogation des connaissances sur le Web des données.

Ce travail offre de nombreuses perspectives : (i) en lexicographie Sens-Texte, (ii) en TALN où le besoin de formaliser les définitions lexicographiques a été noté par Wanner (2003), et (iii) en ingénierie des connaissances, car le formalisme des Graphes d'Unités pourrait apparaître le plus adapté pour représenter les connaissances d'un autre domaine de spécialité que la linguistique.

1. ontolex - Ontology-Lexical Community Group - <http://www.w3.org/community/ontolex/>

Références

- Altman, J. et Polguère, A. (2003). La BDéf : base de définitions dérivée du Dictionnaire explicatif et combinatoire. *In Proceedings of the First International Conference on Meaning-Text Theory (MTT'2003)*, pages 43–54, Paris, France.
- Apresian, J., Boguslavsky, I., Iomdin, L., Lazursky, A., Sannikov, V., Sizov, V. et Tsinman, L. (2003). ETAP-3 linguistic processor : a full-fledged NLP implementation of the MTT. *Proceedings of the First International Conference on Meaning-Text Theory (MTT'2003)*, pages 279–288.
- Bachimont, B. (2000). Engagement sémantique et engagement ontologique : conception et réalisation d'ontologies en ingénierie des connaissances. In Charlet, J., Zaclad, M., Kassel, G. et Bourigault, D., éditeurs : *Ingénierie des connaissances, évolutions récentes et nouveaux défis*, chapitre 19. Eyrolles.
- Barque, L. (2008). *Description et formalisation de la polysémie régulière du français*. Thèse de doctorat, Université Paris 7.
- Barque, L. et Polguère, A. (2009). Structuration et balisage sémantique des définitions du Trésor de la Langue Française informatisé (TLFi). *In Proceedings of the Fourth International Conference on Meaning-Text Theory (MTT'2009)*.
- Flouris, G., Manakanatas, D., Kondylakis, H., Plexousakis, D. et Antoniou, G. (2008). Ontology change : classification and survey. *The Knowledge Engineering Review*, 23(02):117–152.
- Gruber, T. (1995). Toward principles for the design of ontologies used for knowledge sharing ? *International journal of human-computer studies*, 43(5-6):907–928.
- Gugert, R. (2013). Scénarisation d'interactions avec les objets du formalisme des Graphes d'Unités et prototypage d'un éditeur de définitions lexicographiques formelles. Rapport technique, Inria.
- Iordanskaja, L. et Mel'čuk, I. A. (1990). Semantics of Two Emotion Verbs in Russian : BOJAT'SJA 'to be afraid' & NADEJAT'SJA 'to hope'. *Australian Journal of linguistics*, 10(2):307–357.
- Kahane, S. (2003). The Meaning-Text Theory. *Dependency and Valency, An International Handbooks of Contemporary Research*, 25(1):546–569.
- Leclère, M. (1998). Raisonner avec des définitions de types dans le modèle des graphes conceptuels. *Revue d'intelligence artificielle*, 12(2):243–278.

- Lefrançois, M. (2014). *Représentation des connaissances sémantiques lexicales de la Théorie Sens-Texte : Conceptualisation, représentation, et opérationnalisation des définitions lexicographiques*. Thèse de doctorat, Université de Nice - Sophia Antipolis.
- Lefrançois, M. et Gandon, F. (2011). ILexicOn : Toward an ECD-Compliant Interlingual Lexical Ontology Described with Semantic Web Formalisms. In Boguslavsky, I. et Wanner, L., éditeurs : *Proceedings of the 5th International Conference on Meaning-Text Theory (MTT'2011)*, pages 155–164, Barcelona, Spain. INALCO.
- Lefrançois, M. et Gandon, F. (2013a). 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)*, pages 382–388.
- Lefrançois, M. et Gandon, F. (2013b). Reasoning with Dependency Structures and Lexicographic Definitions using Unit Graphs. In *Proc. of the 2nd International Conference on Dependency Linguistics (Depling'2013)*, Prague, Czech Republic. ACL Anthology.
- Lefrançois, M. et Gandon, F. (2013c). The Unit Graphs Framework : A graph-based Knowledge Representation Formalism designed for the Meaning-Text Theory. In *Proceedings of the 6th International Conference on Meaning-Text Theory (MTT'2013)*, Prague, Czech Republic.
- Lefrançois, M. et Gandon, F. (2013d). The Unit Graphs Framework : Foundational Concepts and Semantic Consequence. In *Proceedings of the 9th International Conference on Recent Advances in Natural Language Processing (RANLP 2013)*, Hissar, Bulgaria. ACL Anthology.
- Lefrançois, M., Gandon, F., Giboin, A. et Gugert, R. (2014). Un éditeur de définitions formelles pour les connaissances lexicales de la théorie Sens-Texte. In *Proc. 25e Journées Francophones d'Ingénierie des Connaissances (IC'2014)*, Clermont Ferrand, France.
- Lefrançois, M., Gugert, R., Gandon, F. et 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. et 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):1–66.
- Mel'čuk, I., Clas, A. et Polguère, A. (1995). *Introduction à la lexicologie explicative et combinatoire*. Duculot, Paris/Louvain-la-Neuve.
- Mel'čuk, I. A. et Wanner, L. (2001). Towards a lexicographic approach to lexical transfer in machine translation (illustrated by the German–Russian language pair). *Machine Translation*.
- Polguère, A. (1990). *Structuration et mise en jeu procédurale d'un modèle linguistique déclaratif dans un cadre de génération de texte*. Thèse de doctorat, Département de linguistique, Université de Montréal.
- Star, S. L. et Griesemer, J. R. (1989). Institutional Ecology, 'Translations' and Boundary Objects : Amateurs and Professionals in Berkeley's Museum of Vertebrate Zoology, 1907-39. *Social Studies of Science*, 19(3):387–420.
- Tesnière, L. (1959). *Éléments de syntaxe structurale*. C. Klincksieck (Colombes, Impr. ITE).
- 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)*, pages 16–28.
- Wanner, L. (2009). Explanatory Combinatorial Lexicology and NLP Applications in Specialized Discourse. In *First International Workshop on Terminology and Lexical Semantics (TLS'09)*, pages 54–66.

Summary

The Meaning-Text Theory (MTT) is one of the most precise and exhaustive linguistic theories. Representing knowledge of the MTT is a challenge for applications in lexicography, but also for injecting fine grained linguistic knowledge in NLP systems. This paper describes the methodology and the main outcomes of our research in conducting a Knowledge Engineering approach to model the lexical knowledge of the MTT and formalize its semantics. We focus more specifically on linguistic predicates, linguistic representations, and lexicographic definitions. We drew our inspiration from Bachimont's methodology, and operated in three main steps : (i) we extended the domain conceptualization, (ii) we chose an adequate Knowledge Representation formalism, and (iii) we operationalized the formalism for logical reasoning, and for sharing and querying knowledge on the web of data.