

Les Outils Informatiques au Service de la Terminologie Saussurienne

Ruimy, Nilda, Piccini, Silvia & Giovannetti, Emiliano

Istituto di Linguistica Computazionale *Antonio Zampolli*

Consiglio Nazionale delle Ricerche

Pise - Italie

{nilda.ruimy silvia.piccini et emiliano.giovannetti}@ilc.cnr.it

1 Introduction

Ferdinand de Saussure (1857-1913) est universellement reconnu comme le père fondateur de la linguistique moderne. Sa pensée et son œuvre ont eu un impact considérable dans le domaine des sciences humaines et sociales et en particulier en linguistique et en philosophie du langage. Un trait remarquable de la personnalité de cet éminent savant est en effet l'éclectisme de sa curiosité intellectuelle. La multiplicité de ses intérêts l'a amené à conduire d'importantes recherches dans de nombreux et divers champs du savoir. Néanmoins, Ferdinand de Saussure n'a jamais publié ses écrits sur des thèmes de linguistique générale, et bien que son œuvre ait été largement diffusée à travers le monde, sa pensée a été en grande partie reconstruite et interprétée par ses étudiants et ses disciples. L'ouvrage qui l'a rendu célèbre, le *Cours de Linguistique Générale*, fut en effet rédigé à partir de notes prises par certains de ses étudiants durant les trois cours de linguistique générale que Saussure tint à l'Université de Genève, de 1907 à 1911. Le livre fut publié à titre posthume, en 1916, par deux de ses disciples et collègues, Charles Bally et Albert Sechehaye qui élaborèrent ces notes, non sans y ajouter des commentaires personnels pouvant parfois détourner le lecteur de l'enseignement saussurien. Ces interventions éditoriales de Bally et Sechehaye furent mises en évidence dans l'ouvrage exégétique de Robert Godel *Les sources manuscrites* (1957), ainsi que dans celui de Rudolf Engler *Cours de linguistique générale* (1968-1974) qui mit en parallèle le CLG, les notes des étudiants et les écrits originaux de Saussure.

En réalité, la pensée saussurienne est beaucoup plus subtile et moins catégorique qu'elle ne paraît dans le Cours: d'une part, Saussure confesse souvent ses doutes sur des points cruciaux et fait de ces doutes mêmes son heuristique; d'autre part, il lutte contre l'absence d'une réflexion épistémologique de la linguistique de son époque et propose parfois même une relecture des concepts fondamentaux. Ces deux pôles apparaissent clairement dans les notes et les écrits autographes du maître genevois. C'est pourquoi, afin de saisir pleinement la pensée de Saussure, il est nécessaire de fonder sa recherche sur des écrits authentiques.

Concernant ces écrits, Charles Bally et Léopold Gautier publièrent en 1922 le *Recueil des publications scientifiques de Ferdinand de Saussure* contenant des manuscrits dont la version finale avait été validée par l'auteur lui-même. En 2002, sous le titre de *Écrits de linguistique générale*, Rudolf Engler et Simon Bouquet publièrent des documents autographes découverts en 1996 dans l'orangerie de la maison familiale de Saussure, à Genève. Parmi eux figurait une première rédaction manuscrite, intitulée *De l'essence double du langage*, du livre sur la linguistique générale auquel le linguiste genevois avait souvent fait allusion, mais qui semblait définitivement perdue.

Bien que fragmentaires, les écrits authentiques de Saussure revêtent une importance fondamentale car ils révèlent la complexité du système philosophique et sémiologique élaboré par leur auteur. Ils mettent également en lumière l'attention toute particulière que Saussure accorde à la terminologie dont il déplore le manque de rigueur dans le domaine linguistique, à son époque. Lui-même utilise souvent un vocabulaire très particulier afin d'exprimer des idées et des thèses qui ont eu une portée révolutionnaire dans le domaine de la linguistique. Il forge des néologismes ou confère un sens nouveau à des mots existants, il emploie quelques termes de façon éphémère, change la dénotation de certains concepts au fil du temps, créant ainsi une terminologie qui lui est propre et qui, de nos jours encore, caractérise fortement sa pensée et la contribution fondamentale qu'il a apportée à la linguistique moderne.

La terminologie de Saussure a fait l'objet de plusieurs études. La documentation la plus complète en la matière remonte cependant à 1968. Il s'agit d'une publication de 57 pages intitulée *Lexique de la terminologie saussurienne*, rédigée par Rudolf Engler. Cet opuscule fournit les définitions de 441 termes, assorties de citations mais aussi de commentaires qui, selon Cristina Vallini (2006) pourraient fourvoyer le lecteur. Près d'un demi-siècle plus tard, et à la lumière des manuscrits de l'orangerie, il semble donc opportun d'actualiser l'étude de la terminologie saussurienne.

Le travail que nous nous proposons d'illustrer dans cet article s'inscrit dans le cadre d'un Projet de Recherche d'Intérêt National (PRIN) financé par le gouvernement italien et dirigé par le Professeur Daniele Gambarara. Ce projet, intitulé « *Per un'edizione digitale dei manoscritti di Ferdinand de Saussure* »¹, a pour objectif la création d'un prototype d'édition numérique des textes de F. de Saussure, réalisé à partir d'une sélection de ses manuscrits. A cet effet, chaque manuscrit est numérisé, classé et doté d'un lien hypertextuel qui renvoie à sa transcription ainsi qu'à sa traduction italienne. Un système de gestion de textes permet de consulter, d'annoter et d'accomplir nombre d'études philologiques et critiques sur le corpus de documents numérisés. Les concordances par forme et par lemme produites pour l'ensemble des textes fournissent, entre autres, une sélection de nouveaux termes spécifiques dont la sémantique est décrite dans un thésaurus-lexique.

2 Un nouveau thésaurus

Un des aspects novateurs de ce projet consiste précisément dans la création du premier thésaurus-lexique électronique de la terminologie linguistique saussurienne. L'emploi du terme 'thésaurus-lexique' s'explique par le fait que, comme nous le verrons, la base de connaissances en cours de réalisation offre infiniment plus d'informations qu'un thésaurus traditionnel. La population de ce lexique est constituée par une nomenclature mise à jour, comprenant tous les termes proposés par Godel (1957) et Engler (1968) ainsi qu'un ensemble de mots-clés extraits des *Écrits de linguistique générale*. La nouvelle terminologie extraite des manuscrits actuellement à l'étude y sera également intégrée. À travers ce lexique, nous nous proposons de fournir une représentation structurée de la terminologie saussurienne, de définir le contenu sémantique de chacun des termes ainsi que la nature et l'importance des relations qui les unissent.

De nos jours, l'application des technologies informatiques dans les sciences humaines nous permet d'envisager cette tâche d'une façon bien différente que ne l'a été la compilation manuelle du lexique, réalisée par Engler. À partir des années 70, un nouveau paradigme de lexiques et thésaurus s'est en effet progressivement développé grâce aux instruments et aux méthodes informatiques qui permettent d'acquérir, de mémoriser, traiter et éditer des quantités de données lexicales inconcevables pour la constitution manuelle d'un thésaurus ou d'un dictionnaire.

Un thésaurus est généralement défini comme un répertoire alphabétique de termes reliés par des relations hiérarchiques, associatives et synonymiques. Ces termes constituent le lexique d'un domaine de la connaissance et sont utilisés pour l'indexation des textes et la recherche d'information dans ce domaine. Quelques thésaurus électroniques ont été créés, parmi lesquels il convient de signaler le *Thesaurus mediae et recentioris latinitatis*, élaboré à Rome à l'Istituto del Lessico Intellettuale Europeo e Storia delle Idee du Consiglio Nazionale delle Ricerche, et qui est consacré au lexique philosophique des XVII et XVIII siècles et à la liste des termes du haut Moyen-Âge. Le *Thesaurus Diplomaticus*, une base de données de textes et images de documents diplomatiques en latin, ainsi que le *Thesaurus formarum totius latinitatis a Plauto usque ad saeculum XXum* contenant des formes latines annotées, ont été tous deux réalisés à l'Université Catholique de Louvain. Le *Thesaurus Linguae Graecae*, développé à la California University, à Irvine, contient, quant à lui, la quasi totalité des textes grecs du VII siècle av. J.-C. à l'an 1453 apr. J.-C.

Les thésaurus peuvent s'appuyer sur un système de classification unidimensionnelle ou multidimensionnelle. Ces derniers, appelés thésaurus à facettes, sont plus flexibles que ceux à structure monohiérarchique puisqu'ils fournissent une multiplicité de points d'accès à l'information et donc plus de possibilités de recherches sur les données saisies. Notre but étant, ici, de développer un thésaurus-lexique, nous nous sommes tournés vers l'étude de modèles lexicaux.

Durant les quinze dernières années, de remarquables modèles lexicaux ont été conçus pour le développement de grandes ressources lexicales informatisées. Certains de ces modèles, tels que Wordnet (Fellbaum, 1998), EuroWordNet (Vossen, 1999), et ItalWordNet (Roventini et al., 2003) sont basés sur des réseaux conceptuels ; d'autres comme par exemple FrameNet (Fillmore et al., 2003) et Pattern Dictionary (Hanks et Pustejovsky, 2005) sur des patrons sémantiques ; d'autres encore conjuguent classification ontologique et description de la valence syntaxique et sémantique, c'est le cas du modèle SIMPLE (Lenci et al., 2000) et de Brandeis Semantic Ontology (Pustejovsky et al., 2006).

3 Le choix d'un modèle lexical

Utiliser un modèle existant en l'adaptant aux exigences d'un domaine de la connaissance est, sans aucun doute, l'option la plus avantageuse en matière de coût et de temps mais aussi et surtout en matière de qualité des résultats. Après un examen attentif des différents types de modèles susnommés, nous avons opté pour le modèle sémantique SIMPLE en raison de ses remarquables caractéristiques, et l'avons customisé en vue du développement du thésaurus-lexique électronique de la terminologie saussurienne.

3.1 Le modèle de base

Le modèle lexical SIMPLE fut adopté consensuellement au plan européen en 1998 en vue de créer douze lexiques sémantiques monolingues harmonisés, pour autant de langues européennes. Dès sa conception, il s'est imposé comme standard *de facto* et a ainsi inspiré de manière déterminante le standard international ISO pour les lexiques du Traitement Automatique des Langues, le métamodèle Lexical Markup Framework² (Francopoulo, 2007). Dans le panorama de la Lexicographie Computationnelle, le modèle SIMPLE se distingue par certains aspects particulièrement novateurs.

L'approche théorique qui sous-tend la représentation de l'information sémantique est fortement inspirée des principes fondamentaux de la théorie du Lexique Génératif (Pustejovsky, 1995 ; 2001). L'architecture

flexible du modèle ainsi que la méthodologie de construction du lexique permettent d'encoder dans les entrées lexicales une vaste gamme d'informations hautement structurées couvrant les principaux aspects de la sémantique lexicale des mots ; ceci de manière cohérente, et au degré de granularité souhaité. La structuration conceptuelle du lexique est confiée à une ontologie linguistique mettant en relief le caractère multidimensionnel du sens lexical. L'ontologie SIMPLE est formée de types sémantiques indépendants des langues et a été conçue en prévision d'expansions ou de spécialisations visant à satisfaire aux exigences de domaines d'intérêt particuliers.

Outre la classification ontologique et un certain nombre de traits sémantiques, les unités lexicales sont caractérisées par un réseau de relations sémantiques qui sont en majeure partie le fruit d'une relecture de la Structure des *Qualia* (*Qualia Structure*) que Pustejovsky emprunte à Moravcsik (1975)³. La structure des *qualia*, qui est un des quatre niveaux de représentation sémantique⁴ dans la théorie du Lexique Génératif, est composée de quatre rôles. Ceux-ci permettent respectivement de situer l'entité dénotée parmi d'autres (rôle formel), d'en indiquer la fonction (rôle téléique), l'origine (rôle agentif) et la composition (rôle constitutif). La structure des *qualia* fournit donc un langage formel pour modéliser l'aspect componentiel du sens d'un mot. Les relations *qualia* permettent de représenter l'information concernant les différentes composantes sémantiques qui concourent à définir la structure interne d'une unité lexicale — quelle que soit sa catégorie syntaxique. Elles explicitent ses liens paradigmatiques : hyperonymiques, hyponymiques, méronymiques et holonymiques mais aussi, sur l'axe syntagmatique, les liens aux différents prédicats qui contribuent à éclairer son sens ; ces derniers renseignent en particulier sur l'origine et la fonction de l'entité dénotée. Outre les relations *qualia*, d'autres liens tels que la synonymie, la polysémie logique et la dérivation morphologique sont également définis. Les lexèmes prédicatifs - verbes, noms prédicatifs et adjectifs - sont de surcroît caractérisés par le scénario sémantique dans lequel ils sont impliqués. Celui-ci est décrit en termes de classe aspectuelle et de structure argumentale du prédicat, avec indication du rôle sémantique et des restrictions de sélection sémantique de chacun des arguments.

4 La customisation du modèle : de SIMPLE à SIMPLE_FdS

Le modèle SIMPLE offre donc, pour la structuration du lexique, un cadre riche mais cependant flexible, dans lequel de nombreux éléments sont paramétrables. L'ontologie a été conçue tant de manière à permettre l'élimination des types sémantiques les plus spécifiques figurant au bas de la hiérarchie, qu'en vue d'une expansion globale ou sectorielle. La structure *qualia* se prête aisément à une réduction, une extension ou une customisation des relations, afin de mieux saisir l'organisation conceptuelle d'un domaine particulier ou de répondre aux besoins spécifiques d'une application. L'introduction de traits sémantiques ainsi que de relations non *qualia* supplémentaires est, quant à elle, autorisée.

4.1 L'ontologie de domaine

La première phase du processus de customisation du modèle de base s'est focalisée sur la conception d'une ontologie lexicale de domaine permettant de structurer la terminologie saussurienne. Pour l'usage du terme 'ontologie', nous nous référons à la définition la plus répandue en linguistique et en informatique : «An ontology is an explicit specification of a conceptualization. [...] An ontology defines (specifies) the concepts, relationships, and other distinctions that are relevant for modelling a domain.»⁵ (Gruber, 1993 : 200).

Construire une ontologie, c'est modéliser et représenter le schéma conceptuel d'un domaine de la connaissance, en vue de partager une représentation formelle de ses éléments constitutifs et des relations qu'ils entretiennent. En l'occurrence, notre source d'inspiration étant le système de types sémantiques du modèle SIMPLE, l'ontologie de domaine a été réalisée conformément à l'architecture et aux principes de conception de notre archétype. Celui-ci reflétant un précepte fondamental de la théorie du Lexique Génératif selon lequel les unités lexicales sont des entités multidimensionnelles qui diffèrent en degré de complexité véhiculé par leur sémantique lexicale, l'ontologie de domaine a donc été, elle aussi, conçue de façon à représenter ces différences de complexité interne du sens des mots.

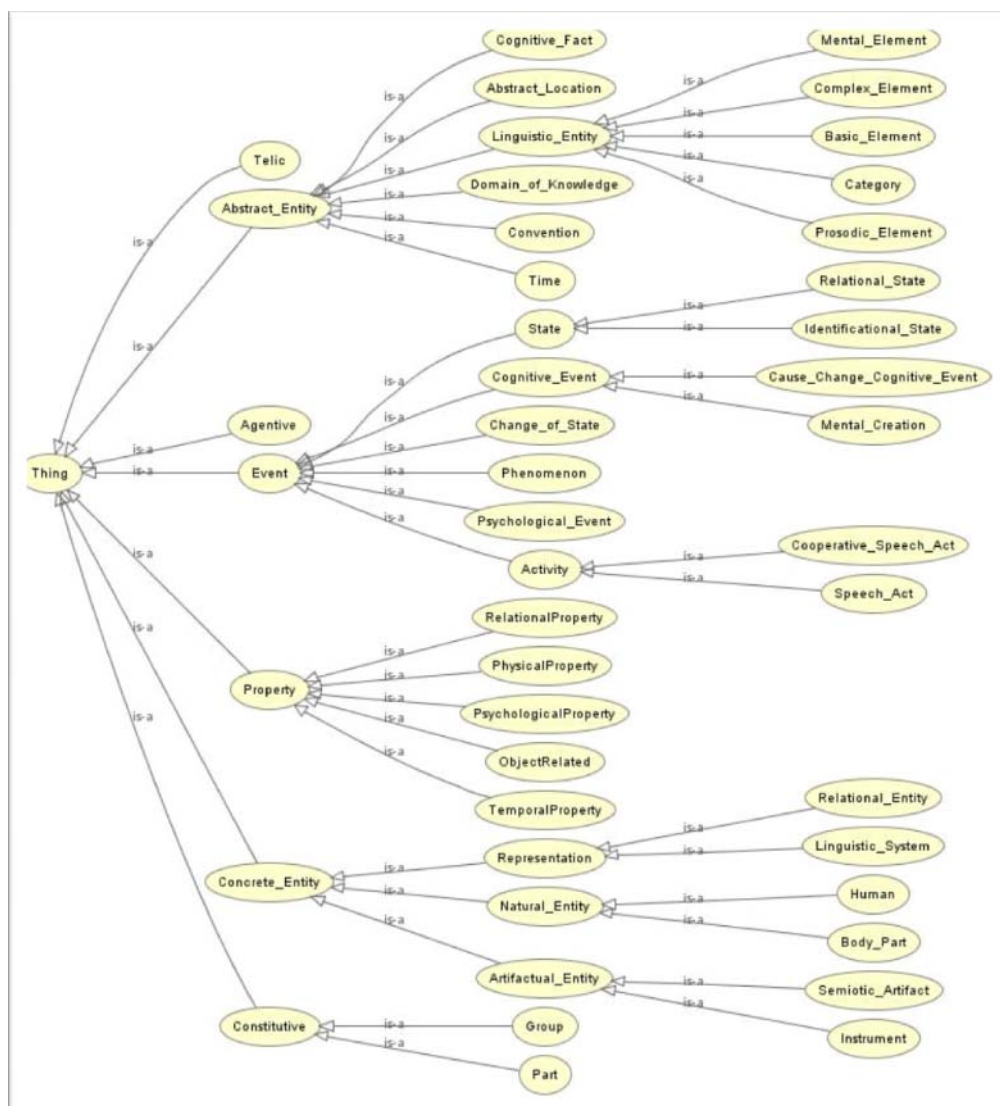
4.1.1 Identification des classes conceptuelles

La première étape dans la création de l'ontologie *SIMPLE_FdS* a été d'identifier un ensemble de classes sémantiques à partir de leurs instances. L'extraction de nouveaux termes des manuscrits à l'étude n'étant pas achevée, nous avons analysé les termes recueillis par Godel et Engler - termes indubitablement fondamentaux dans la pensée saussurienne. Chaque sens d'un terme a reçu un classement sémantique préliminaire et parfois même approximatif. De cette classification ont émergé les classes conceptuelles les plus pertinentes de la future ontologie. Ces concepts, souvent situés à un niveau intermédiaire de l'ontologie, ont ensuite été définis et organisés en une hiérarchie. Puis, dans un processus toujours en cours, ils ont été généralisés ou spécialisés à travers la création de classes sémantiques plus vastes ou de sous-classes, en combinant les approches ascendante et descendante. À cet égard, nous avons pris soin d'éviter tant une trop grande généralisation que des distinctions trop fines et qui se révèlent souvent peu utiles.

L'étape successive a été consacrée à la définition de la structure interne des classes (ou types) sémantiques : a) leurs propriétés (attributs et valeurs), en distinguant celles qui sont communes à toutes les instances d'un type sémantique et qui en sont donc un trait définitoire (attribut obligatoire) de celles qui ne caractérisent qu'un sous-ensemble des termes appartenant à un type (attribut autorisé); b) les valeurs possibles des attributs : valeur booléenne, caractères, liste, chiffre, ainsi que leur cardinalité (le nombre minimal et maximal de valeurs distinctes autorisées).

4.1.2 La structuration des concepts

En accord avec notre modèle d'ontologie, le système de types sémantiques *SIMPLE_FdS* est fondé sur des relations conceptuelles tant hiérarchiques que non hiérarchiques, selon le principe de l'héritage orthogonal⁶ (Pustejovsky et Boguraev, 1993). Il est donc constitué de types unidimensionnels ainsi que de types multidimensionnels. Les premiers sont pleinement caractérisables par une relation hyperonymique ; ainsi, le type sémantique 'Natural_Entity' est correctement défini comme un sous-type de 'Concrete_Entity'. Les types multidimensionnels traduisent, quant à eux, la multidimensionnalité du sens à travers la combinaison d'une relation de subsomption et la référence à d'autres dimensions de sens. Ces dimensions de sens sont exprimées au moyen de la structure des *qualia*, qui est inspirée de la notion aristotélicienne de modes d'explication (*Aitiai*) ; ainsi, le type sémantique 'Artifactual_Entity' ne peut être défini que comme un sous-type de 'Concrete_Entity' englobant également une dimension agentive et une dimension téléique. Dans le cadre de l'ontologie, les rôles *qualia* définissent les propriétés distinctives des types sémantiques et différencient leur constitution sémantique interne.



Graphique 1 : Visualisation de l'ontologie *SIMPLE_FdS* (version 0.2) avec le logiciel Protégé

Il n'est pas surprenant de constater que dans l'ontologie de domaine, les types sémantiques les plus élevés dans la hiérarchie, qui sont les plus génériques et consensuels puisqu'ils expriment des concepts très généraux, reflètent en grande partie ceux de l'ontologie SIMPLE. Au plus haut niveau de l'ontologie de domaine figurent en effet – outre les types Constitutif, Agentif et Télique qui permettent de fournir une représentation sémantique d'entités dont le rôle formel est indéfinissable et qui n'instancient lexicalement que le rôle Constitutif, Agentif ou Télique, ex. : *ensemble*, *origine du langage* ou *objet linguistique* – les types *Concrete_Entity*, *Abstract_Entity*, *Property* et *Event*. De même, une portion de l'ontologie générale a inspiré la construction de hiérarchies de types pour la description du métalangage linguistique.

Cette ontologie de domaine, constituée de 47 types sémantiques, a été conçue tant dans le respect des principes généraux de construction d'une ontologie que dans celui des contraintes spécifiques imposées par le modèle lexical adopté. Nous nous sommes efforcés d'adhérer aux principes de complétude et de

cohérence et de prévoir la possibilité d'extensions ou raffinements. Notre objectif principal a été de tendre vers une modélisation la plus consensuelle possible des concepts et des relations constitutives propres à notre domaine de la connaissance, ceci dans le but de permettre une recherche d'information efficace et exhaustive et l'acquisition de connaissances lexicales à partir de textes.

Le réseau de classes sémantiques ainsi conçu ne représente, il est clair, qu'une des modélisations possibles du domaine de la terminologie saussurienne. Le développement d'une ontologie est en effet un processus créatif dans lequel un facteur de subjectivité s'inscrit inévitablement ; en dépit de tout effort de consensus, une ontologie reflète toujours un point de vue. Nous assumons donc pleinement la responsabilité du choix et de la définition des concepts qui caractérisent la connaissance du domaine, de leurs relations, ainsi que du niveau de spécification du schéma conceptuel tout entier. D'autre part, la version actuelle de l'ontologie pourrait être sujette à évolution, et en particulier à des raffinements visant à une définition plus précise des types sémantiques ainsi qu'à des extensions pour l'introduction de nouveaux concepts. Il est en effet possible que, dans une phase plus avancée du développement du lexique, nous soyons confrontés à des instances de concepts non représentés dans l'ontologie. Cette évolution doit, bien évidemment, s'effectuer de manière à ne pas invalider les contraintes du modèle et dans le respect de la cohérence du système tout entier.

L'ontologie *SIMPLE_FdS* a été traduite dans le langage de représentation d'ontologie Web OWL (Web Ontology Language). Disposer d'une version OWL de l'ontologie offre un triple avantage : 1) OWL étant basé sur une sémantique formelle, il est possible d'effectuer un grand nombre de tests sur l'ontologie et sa cohérence (restrictions doubles, parents multiples, etc.) ; 2) il existe une série de plugins de visualisation pour l'éditeur d'ontologies Protégé qui simplifient l'édition et la consultation ; 3) OWL est le standard proposé depuis 2004 par le W3C pour la représentation et le partage d'ontologies sur le Web. L'ontologie représentée dans le graphique 1 a été éditée avec Protégé-OWL, une plate-forme open-source qui fournit une série d'outils pour la création d'ontologies.

4.2 Traits et relations sémantiques

Outre l'ontologie, deux autres composants du modèle *SIMPLE_FdS*, i.e., un groupe de traits sémantiques et un vaste ensemble de relations sémantiques, permettent d'exprimer une large typologie d'informations caractérisant le contenu sémantique des mots.

Les traits sémantiques expriment des informations telles que le domaine d'usage d'une unité lexicale, ex. : 'Semantics', 'Morphology' ; certaines propriétés distinctives transversales à la hiérarchie des types et qui permettent donc de regrouper les termes, ex. : 'Plus_Conventional', quelle que soit leur classification ontologique ; ou encore des dimensions de sens qui sont clairement perçues dans le contenu sémantique d'un mot mais ne peuvent être exprimées du point de vue lexical au moyen d'une relation sémantique 'Agentive_yes', 'Telic_yes'.

Les relations sémantiques jouent, quant à elles, un rôle fondamental dans la modélisation des données lexicales. Elles expriment, sous forme de triplets <unité sémantique source, relation, unité sémantique cible>, une relation binaire entre des acceptions. Elles proviennent en grande partie de la *Extended Qualia Structure* du modèle SIMPLE, qui est elle-même, comme précédemment observé, le fruit d'une relecture de la structure des *qualia*. Les relations de l'*Extended Qualia* permettent de fournir une représentation structurée de la multidimensionnalité d'un sens et de saisir la nature des relations (intra- et extracatégories) existant entre les unités lexicales, à travers la base de données. L'ensemble des

relations dont dispose le modèle *SIMPLE_FdS* permet donc, en résumé, de relier une entité-objet à une entité plus générique, à ses propriétés, son origine, sa fonction, à ses éléments constitutifs ou au tout dont elle fait partie ; une entité-événement, peut, quant à elle, être reliée à son agent, patient, instrument, éventuel lieu d'action, ainsi qu'à son effet ou résultat. D'autres types de relations sont également représentées, parmi lesquelles la synonymie et l'antonymie.

Dans le modèle *SIMPLE_FdS*, les traits et relations sémantiques du modèle de base jugés appropriés pour rendre compte de la terminologie saussurienne ont été retenus. D'autres traits et relations spécifiques ont été créés afin d'exprimer des propriétés et des liens qui caractérisent les termes et plus généralement l'organisation conceptuelle de ce domaine particulier de la connaissance.

Seize nouveaux traits sémantiques ont été introduits, qui expriment de nouveaux domaines d'usage, renseignent sur la source, la fréquence d'occurrence et la période d'attestation d'un terme ainsi que sur le caractère cognitif de l'entité qu'il dénote.

À ce jour, seize relations sémantiques ont été créées dans le but de représenter a) différents types de changements terminologiques (*hasPrevious* / *hasSubsequent* / *hasOther_Denomination*) exprimant soit la volonté de Saussure lui-même (*signifiant hasPreviousDenomination image acoustique*) soit celle de l'éditeur (*chaînon isEdSubstitutedWith fragment*) (*fragment substitutesFor chaînon*) ; b) une relation de quasi-synonymie (*diachronie hasNearSynonym évolution*) ; c) une opposition (*consonante isOpposedto sonante*) ; d) l'appartenance à une sphère spécifique (*sémantique pertainsTo linguistique*), (*métaplasme belongsTo diachronie*) ; e) un trait fondamental de la pensée saussurienne (*parole duality langue*) ; f) différents types d'implication réciproque (*signifiant directlyImplies signifié*), (*image auditive indirectlyImplies idée*) ; g) une implication non réciproque, inspirée de (Cruse, 1986 : 123) (*parasème hasEndonym sème*), dans laquelle le terme conteneur (ici, l'unité sémantique source) englobe le sens du terme représenté par l'unité sémantique cible ; h) un ordre linéaire ou temporel (*frontière syllabique precedes explosion*) (*suffixe follows racine*) ; i) un lieu abstrait (*trésor hasAbstractLocation pensée*).

5 Le niveau terminologique

Le modèle conceptuel une fois établi, la phase successive consiste à représenter formellement la sémantique lexicale des instances des concepts ontologiques. Les moyens descriptifs qui ont été présentés permettent la création d'entrées lexicales (cf. tableaux 1 et 2) particulièrement riches d'informations.

Une entrée sémantique décrit une seule acception d'un lexème simple ou complexe. Le sens analysé y est défini par un ensemble d'informations structurées couvrant un large éventail d'aspects sémantiques : classification ontologique ; domaine d'usage ; définition fournie par Saussure lui-même⁷ et par Engler et Godel ; collocations ; type d'événement dénoté, i.e. état, procès, transition ; traits sémantiques distinctifs particuliers (source, fréquence et période d'attestation d'un terme) ; relation de synonymie et d'antonymie ; relation de dérivation morphologique ; relations de la *Extended Qualia Structure* ; structure argumentale du prédicat avec indication du rôle sémantique ainsi que des restrictions sémantiques de ses arguments. Une attention particulière est donc réservée aux liens existant entre les différentes unités sémantiques. Les relations sémantiques établissent un lien entre le sens décrit et un certain nombre d'événements ou d'entités qui lui sont strictement liés ; leur nature explicite permet, de surcroît, de définir le rôle de ces événements ou entités dans la sémantique lexicale du mot décrit.

<p>id = "USem4514uni-spatialité" Naming = "uni-spatialité"</p> <p>Saussure definition = "Principe de l'uni-spatialité (si l'on considère le sème) ayant pour conséquence dans le sème la divisibilité par tranches (toujours dans le même sens et par coupures identiques) [...] au lieu de la divisibilité pluriforme qu'on aurait par exemple dans un système visuel direct." 3317.2. Il faut s'astreindre à dire uni-spatialité du signe linguistique chaque fois, afin de faire sentir que ce n'est pas un caractère général du sème." 3318.1.</p> <p>Definition = "Linéarité du signifiant [l'opposition entre deux sèmes se règle au moyen de tranches allant dans le même sens et n'arrivant qu'une à une]. "(Engler, 51).</p> <p>Semfeaturelist= "SemanticType Property SuperType Thing Domain LINGUISTICS "</p> <p>Relations: semr = "isa" target = "USemD2841propriété" (Property) semr = "isaPropertyOf" target = "USemD24signifiant" (Concrete_Entity) semr = "isaPropertyOf" target = "USemD4057syntagme" (Complex_Element) semr = "isaPropertyOf" target = "USemD06chaîne_de_la_parole" (Speech_Act) semr = "belongsTo" target = "USemD4058parole" (Speech_Act) semr = "hasSynonym" target = "USemD76linéarité" (Property) semr = "deadjectivalNoun" target = "USemD9224spatial" (Physical_Property)</p> <p>Predicative Representation: predicate = "PREDSpatial#1" typeoflink = "AdjectiveNominalization" Arguments: id = "ARG0spatial#1" semanticrole = "Role_Underspecified" semanticrestriction = " SemanticType Thing"</p>	<p>id = "USemD9066temps_homogène" Naming = "temps_homogène"</p> <p>Saussure definition = "Espace (de temps) acoustiquement homogène, correspondant à l'articulation d'un phonème." 748 (3335,1).</p> <p>Definition = "Correspondance d'un temps acoustique et d'un temps articuloire 749> E 65/64-66/65. Dans le Cours III, Saussure dira moment ("des moments, pas des temps" 725)" (Engler, 50). "Espace homogène dans la chaîne acoustique; acte articuloire qui lui correspond." (Godel, 279).</p> <p>Semfeaturelist= " SemanticType Time SuperType Abstract_Entity Domain LINGUISTICS Domain PHONOLOGY DurationUnderspecified"</p> <p>Relations: semr = "isa" target = "USemD9120espace_temporel" (Time) semr = "concerns" target = "USemD2865chaîne_phonique" (Group) semr = "Hasproperty" target = "USemD9122homogénéité" (Property) semr = "isRelatedTo" target = "USem8phonème" (Basic_Element) semr = "belongsTo" target = "USemD4058parole" (Speech_Act) semr = "pertainsTo" target = "USem14phonologie" (Domain_of_Knowledge) semr = "hasSubsequentDenomination" target = "USemD89121moment" (Time)</p>
--	---

<p>id = "USemD4058parole" Naming = "parole"</p> <p>Definition = "Exécution du langage à travers le code de la langue (le langage moins la langue)." (CLG, Engler, 38). "Tout ce qui est amené sur les lèvres par le besoin du discours et par une opération particulière (IR2.24). Acte de l'individu réalisant sa faculté du langage au moyen de la convention sociale qui est la langue (IIR6-7)." (Godel, 271).</p> <p>Semfeaturelist= "SemanticType Speech_Act SuperType Act Domain LINGUISTICS</p>	<p>id = "USem11morphologie" Naming = "morphologie"</p> <p>Saussure definition = "Science qui traite des unités de son correspondant à une partie de l'idée, et du groupement de ces unités. Le vrai nom de la morphologie serait: la théorie des signes, et non des formes." 3293 &2.</p> <p>Definition = "Dans les subdivisions de la grammaire (au sens traditionnel), la morphologie s'oppose essentiellement à la phonétique, tandis qu'il est difficile de la distinguer de la syntaxe: "Etude des formes et des fonctions, c'est la même chose" 2136 (Engler, 35) (Godel, 269).</p>
---	---

<p>EventType Process IntentionalityYes"</p> <p>Relations:</p> <p>semr = "isa" target = "USemD2892exécution" (Speech_Act)</p> <p>semr = "concerns" target = "USemD4037langage" (Property)</p> <p>semr = "hasInstrument" target = "USemD402langue" (Linguistic_System)</p> <p>semr = "hasProperty" target = "USemD76linéarité" (Property)</p> <p>semr = "isTypicalOf" target = "USemD8383sujet_parlant" (Human)</p> <p>semr = "duality" target = "USemD402langue" (Linguistic_System)</p> <p>semr = "hasSynonym" target = "USemD9223langue_discursive" (Speech_Act)</p> <p>semr = "hasSynonym" target = "USemD33discours" (Speech_Act)</p>	<p>Collocations = "morphologie historique; - rétrospective"</p> <p>Semfeaturelist= "SemanticType Domain_of_Knowledge SuperType Abstract_Entity Domain LINGUISTICS Domain MORPHOLOGY"</p> <p>Relations:</p> <p>semr = "isa" target = "USemD15science" (Abstract_Entity)</p> <p>semr = "concerns" target = "USemD12forme" (Relational_Entity)</p> <p>semr = "concerns" target = "USemD17unité_morphologique" (Basic_Element)</p> <p>semr = "belongsTo" target = "USemD22synchronie" (State)</p> <p>semr = "pertainsTo" target = "USem12grammaire" (Domain_of_Knowledge)</p> <p>semr = "hasPurpose" target = "USemD18groupement" (Cause_Change_Cognitive_Event)</p> <p>semr = "hasOtherDenomination" target = "USemD19théorie_des_signes" (Domain_of_Knowledge)</p>
---	--

Tableaux 1 et 2 - Exemples d'entrées lexicales

Privilégiant les termes déjà répertoriés provenant des écrits authentiques de Saussure, 375 entrées lexicales ont été créées à ce jour, parmi lesquelles 284 noms simples ou complexes, 70 adjectifs et 21 verbes. Ces entrées sont mémorisées dans une base de données gérée par un système qui permet un accès facile et simultané pour la création, le stockage, la consultation et la modification des entrées lexicales. Le système de gestion de la base de connaissances comprend un ensemble de requêtes en SQL (*Structured Query Language*), un langage de programmation déclaratif créé pour le contrôle et la manipulation de données mémorisées dans les bases de données relationnelles. Ces requêtes permettent d'effectuer des contrôles de qualité sur la cohérence formelle des informations mémorisées. Le système supporte en outre des fonctions de requêtes avancées permettant une recherche d'information lexicale rapide, efficace et complète et qui fournit des résultats très fins. À titre d'exemple, une requête concernant l'usage d'une relation sémantique déterminée, ex. : 'resultingFrom' (avec une éventuelle restriction sur l'unité sémantique cible, ex. : *association*) restitue tous les couples de termes liés par cette relation. À travers une requête concernant une unité lexicale, ex. : *signifiant*, l'ensemble des termes qui lui sont associés sont sélectionnés et classés selon le type de relation qu'ils entretiennent avec le mot clé. Une requête portant sur un trait sémantique, ex. 'plus_sound', saisit en revanche toutes les unités sémantiques concernées, indépendamment de leur classification ontologique.

Source Semantic Unit	Naming	Semantic Type	Sem. Relation	Target Semantic Unit
USemD9838analogie	analogie	Mental_Creation	resultingFrom	USemD4039association
USemD9839novation_analogique	novation_analogique	Mental_Creation	resultingFrom	USemD4039association
USemD2854paradigme	paradigme	Relational_State	resultingFrom	USemD4039association
USem4453rapport_associatif	rapport_associatif	Relational_State	resultingFrom	USemD4039association

Tableau 3 - Requête portant sur une relation sémantique spécifique associée à un mot cible déterminé

Target Semantic Unit	Semantic Relation	Source Semantic Unit
USemD24signifiant	hasPart	USemD31onymique
USemD24signifiant	hasPart	USemD36sème
USemD24signifiant	hasPart	USem5signe
USemD24signifiant	hasPart	USemD27signe_linguistique
USemD24signifiant	directlyImplies	USemD35signifié
USemD24signifiant	indirectlyImplies	USemD23sens
USemD24signifiant	indirectlyImplies	USem35pensée
USemD24signifiant	isa	USem4forme
USemD24signifiant	isPartOf	USem8phonème
USemD24signifiant	isPropertyOf	USemD76linéarité
USemD24signifiant	isPropertyOf	USem4514uni-spatialité
USemD24signifiant	hasOtherDenomination	USemD4050image_verbale
USemD24signifiant	hasOtherDenomination	USemD26son
USemD24signifiant	hasOtherDenomination	USemD9010image_auditive
USemD24signifiant	hasOtherDenomination	USemD9007impression_acoustique
USemD24signifiant	hasSubsequentDenomination	USemD34image_acoustique
USemD24signifiant	hasSynonym	USem10signe_linguistique
USemD24signifiant	hasSynonym	USem139sôme
USemD24signifiant	hasSynonym	USemD25apostème

Tableau 4 - Requête sur un mot cible dans l'ensemble des relations auxquelles il est associé

Semantic Unit	Semantic Type	Semantic Feature
USemD2892exécution	Speech_Act	Plus_sound
USemD60008système_phonologique	Group	Plus_sound
USemD60012unité_phonatoire	Basic_Element	Plus_sound
USemD60011unité_phonique	Basic_Element	Plus_sound
USemD2860fragment	Basic_Element	Plus_sound
USem4414sonorité	Constitutive	Plus_sound
USemD2861anneau	Basic_Element	Plus_sound
USemD2885réalisation	Speech_Act	Plus_sound
USem32acte_linguistique	Speech_Act	Plus_sound

USemD33discours	Speech_Act	Plus_sound
USemD4031chaîne_parlée	Speech_Act	Plus_sound
USemD4058parole	Speech_Act	Plus_sound
USemD60006chaîne_de_la_parole	Speech_Act	Plus_sound
USemD60024phonation	Speech_Act	Plus_sound
USemD2863articulation	Speech_Act	Plus_sound
USemD2889chaîne_significative	Speech_Act	Plus_sound
USemD2869circuit_de_la_parole	Speech_Act	Plus_sound
USem4forme	Concrete_entity	Plus_sound
USemD26son	Concrete_entity	Plus_sound
...

Tableau 5 - Requête sur les unités lexicales associées à un trait sémantique

6 Conclusions et perspectives

Le thésaurus-lexique électronique que nous avons présenté est basé sur un modèle lexical qui allie la complexité et la rigueur à des traits innovants. Une ressource lexicale ainsi structurée et formalisée offre de nombreux avantages. En phase de création, elle permet un développement collaboratif rigoureux, même à distance ; elle favorise une représentation systématique de généralisations concernant des phénomènes linguistiques ; enfin, elle assure la cohérence et la complétude de l'information saisie. En matière de maintenance, elle permet de réaliser des contrôles de cohérence formelle de l'information et d'effectuer des modifications ou corrections sur des ensembles de données ; elle offre une grande facilité d'enrichissement de l'information ainsi que d'extension de la couverture lexicale. En ce qui concerne son exploitation, outre ses caractéristiques d'outil de description et de désambiguïsation sémantique très efficace, cette ressource lexicale permet de rechercher ou d'inférer une vaste gamme d'informations. La structuration multidimensionnelle des concepts du domaine de référence ainsi qu'une représentation sémantique hautement structurée, riche, diversifiée et expressive en font un outil de recherche lexicale particulièrement performant. Des requêtes de recherche d'information peuvent en effet être effectuées très simplement et rapidement sur n'importe quel type de données mémorisées, i.e. relations, traits sémantiques ou unités sémantiques, ces données pouvant être interrogées tant singulièrement qu'en combinaison. Effectuées dans un environnement formalisé qui garantit la complétude des résultats obtenus, ce genre de recherche lexicale permet d'acquérir une vision plus approfondie de la nature componentielle et relationnelle du sens des mots.

Le thésaurus-lexique sémantique de la terminologie saussurienne en cours de développement - et qui sera achevé fin 2012 - deviendra, à notre avis, un outil essentiel pour les études saussuriennes. L'organisation structurée de la connaissance lexicale qui y est proposée favorise la recherche sémantique et devrait donc contribuer de manière significative à mieux maîtriser le vocabulaire saussurien et, partant, à éclairer certains aspects originaux de la pensée du grand linguiste genevois.

Les outils informatiques et les méthodologies adoptés pour le développement de ce thésaurus électronique sont entièrement portables et sont destinés à être utilisés dans des domaines similaires pour étudier l'œuvre d'autres écrivains, philosophes et savants. En fait, dans le cadre d'un ERC Advanced Project⁸, un

thesaurus électronique de termes représentatifs extraits de la pseudo-Théologie d'Aristote est actuellement en cours de développement. Il est fondé sur une version customisée de notre modèle de base.

Références bibliographiques

- Cabré, M.T. (1999). *Terminology. Theory, methods and applications*. Sager, J.C. (éd.). Amsterdam/Philadelphia : John Benjamins Publishing Company.
- Cruse, D. A. (1986). *Lexical Semantics*. Cambridge : Cambridge University Press.
- Engler, R. (1968). *Lexique de la terminologie saussurienne*. Comité international permanent des linguistes. Publication de la commission de terminologie. Utrecht-Anvers : Spectrum.
- Fellbaum, C. (1998). *WordNet. An Electronic Lexical Database*. Cambridge, MA : The MIT Press.
- Fillmore, C.J. et al. (2003). Background to FrameNet. *International Journal of Lexicography*, 16 (3), 235-250.
- Francopoulo, G. et al. (2007). Lexical Markup Framework: ISO standard for semantic information in NLP lexicons. Proceedings of Lexical-semantic and Ontological Resources Workshop at Gesellschaft für linguistische Datenverarbeitung, Biennial Spring Conference, Tübingen.
- Godel, R. (1957). *Les sources manuscrites du Cours de linguistique générale de Ferdinand de Saussure*. Genève : Droz.
- Gruber, T. R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2), 199-220.
- Guarino, N. (1998). Some Ontological Principles for Designing Upper Level Lexical Resources. Proceedings of the First International Conference on Language resources and Evaluation, Granada.
- Guarino, N., Oberle, D., Staab, S. (2009). What Is an Ontology? In S. Staab and R. Studer (éds.) *Handbook on Ontologies*, Second Edition, International Handbooks on Information Systems, Springer Verlag: 1-17.
- Hanks, P., Pustejovsky, J. (2005). A pattern dictionary for natural language processing. *Revue Française de Linguistique Appliquée* 10(2), 63-82.
- Lenci, A., F. Busa, N. Ruimy, E. Gola, M. Monachini, N. Calzolari, A. Zampolli et al. (2000). SIMPLE Linguistic Specifications LE4-8346 SIMPLE, Deliv. D2.1 & D2.2. Pisa : ILC and University of Pisa.
- Lenci, A. et al. (2000). SIMPLE: A General Framework for the development of Multilingual Lexicons. *International Journal of Lexicography, special issue, Dictionaries, Thesauri and Lexical-Semantic Relations* 13(4), 249-263.
- Moravcsik, J. M. (1975). Aitia as Generative Factor in Aristotle's Philosophy. *Dialogue*, 14, 622-36.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA. : The MIT Press.
- Pustejovsky, J. (2001). Type Construction and the Logic of Concepts. In Bouillon, P. et Busa, F. (éds.), *The Language of Word Meaning*, Cambridge, MA : Cambridge University Press, 91-123.
- Pustejovsky, J. et al. (2006). Towards a Generative Lexical Resource: The Brandeis Semantic Ontology. CD-ROM Proceedings Fifth International Conference on Language resources and Evaluation, Genova.
- Pustejovsky, J., Boguraev, B. (1993). Lexical Knowledge Representation and Natural Language Processing, *Artificial Intelligence, LXIII*, 193-223.

- Rosati, L. (2003). La classificazione a faccette fra Knowledge Management e Information Architecture, itconsult.it. http://www.itconsult.it/knowledge/articoli/pdf/itc_rosati_faccette_e_KM.pdf
- Roventini A. et al. (2003). ItalWordNet: Building a Large Semantic Database for the Automatic Treatment of Italian. In Zampolli et al. (éds.), *Computational Linguistics in Pisa, Special Issue, Vol. XVIII-XIX, Tomo II*, Pisa-Roma : IEPI, 745-792.
- Ruimy, N. et al. (2003). A computational semantic lexicon of Italian: SIMPLE. In Zampolli et al. (éds.), *Computational Linguistics in Pisa, Special Issue, Vol. XVIII-XIX, Tomo II*, Pisa-Roma : IEPI, 821-864.
- Saussure, F. de (1916). *Cours de linguistique générale*. Bally, C. et Sechehaye, A. (éds). Lausanne-Paris : Payot.
- Saussure, F. de (1968-1974). *Cours de linguistique générale*. Engler, R. (éd.). Vol. 1 et 2. Wiesbaden : Otto Harrassowitz.
- Saussure, F. de (1922). *Recueil des publications scientifiques de Ferdinand de Saussure*. In Bally, Ch. et Gautier, L. (éds.). Lausanne : Payot.
- Saussure, F. de (2002). *Écrits de linguistique générale*. Bouquet, S. et Engler, R. (éds.). Paris : Gallimard.
- Sowa, J. F. (2002). Building, Sharing, and Merging Ontologies. <http://www.jfsowa.com/ontology/ontoshar.htm>
- Vallini, C. (2006). Il metalinguaggio di Saussure fra vecchie e nuove scienze. In Gensini, S. e Martone, A. (éds.), *Il linguaggio: Teoria e storia delle teorie, in onore di Lia Formigari*, Napoli : Liguori Editore.
- Vossen, P. (éd.) (1999). *Eurowordnet. A multilingual database with lexical semantic networks*. Dordrecht : Kluwer.

¹ « Pour une édition numérique des manuscrits de Ferdinand de Saussure ».

² ISO-24613:2008.

³ Moravcsik, quant à lui, fournit une interprétation des quatre causes aristotéliennes .

⁴ Les trois autres niveaux étant: la structure argumentale, la structure événementielle et la structure d'héritage lexical.

⁵ Une ontologie est une spécification explicite d'une conceptualisation. [...] Une ontologie définit (spécifie) les concepts, les relations ainsi que d'autres distinctions pertinentes pour la représentation d'un domaine.

⁶ Selon ce principe, les différentes facettes du sens d'une unité lexicale peuvent être héritées de nœuds différents mais une propriété donnée est héritée d'un *unique* parent.

⁷ Dans les écrits autographes de Saussure, la définition d'un terme n'est pas toujours donnée de manière explicite. Dans ce cas, seules les définitions fournies par Engler et Godel sont reportées dans l'entrée lexicale.

⁸ ERC Ideas Advanced Grant 249431 *Greek into Arabic. Philosophical Concepts and Linguistic Bridges*.