

# Le lexique électronique de la terminologie de Ferdinand de Saussure : une première

## Section 16

Silvia Piccini, Emiliano Giovannetti, Nilda Ruimy

Le Projet de Recherche d'Intérêt National financé par le gouvernement italien et intitulé « *Per un'edizione digitale dei manoscritti di Ferdinand de Saussure* »<sup>1</sup> a pour objectif la création d'un prototype d'édition numérique des textes du grand linguiste genevois, réalisé à partir d'une sélection de ses documents autographes. A cet effet, chaque manuscrit est numérisé, classé et doté d'un lien hypertextuel renvoyant à sa transcription. 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, un ensemble de termes caractéristiques dont la sémantique est décrite. Dans ce projet, l'un des volets novateurs que nous présentons ici est en effet la création du premier thésaurus-lexique électronique de la terminologie linguistique saussurienne.

Afin de réaliser cette base de connaissance, nous nous sommes inspirés du modèle lexical *SIMPLE* (Lenci, 2000a,b) qui a permis le développement de vastes lexiques sémantiques informatisés (Ruimy, 2003), et l'avons adapté aux exigences de notre domaine de la connaissance. Ce choix nous a paru le plus judicieux, tant en matière de coût et de temps qu'en matière de qualité des résultats. Le lexique réalisé a été conçu dans le but d'offrir un outil performant pour les études saussuriennes. Il fournit en effet une représentation structurée de la terminologie de Saussure, une définition du contenu sémantique de chacun des termes ainsi qu'un cadre explicite de la nature et de l'importance des liens qui les unissent. Ce faisceau d'informations devrait donc contribuer de manière significative à mieux maîtriser le vocabulaire saussurien et, partant, à éclairer certains aspects originaux de la pensée du père de la linguistique moderne.

La structuration du lexique a nécessité la création d'une ontologie lexicale. Pour ce faire, une approche centrifuge a été adoptée, à savoir l'identification des concepts centraux du domaine d'intérêt puis, tour à tour, leur généralisation et leur spécialisation. L'ontologie ainsi modélisée est composée de 43 concepts hiérarchisés, avec une profondeur de 4 niveaux. Certaines classes sémantiques sont en outre restreintes au moyen de traits et/ou de relations sémantiques définitionnelles. Quelque 58 types de relations lexico-sémantiques permettent de relier les instances de l'ensemble des classes ontologiques, tandis que 32 types de traits encodent des informations caractérisant soit une classe toute entière soit une instance spécifique.

La sémantique lexicale de chacune des instances d'un concept ontologique est représentée dans une entrée lexicale. Dans chaque entrée, une et une seule acception d'un terme, simple ou complexe, est très richement définie au moyen d'un vaste ensemble d'informations formalisées et hautement structurées, couvrant un large éventail d'aspects sémantiques. Le sens analysé est tout d'abord associé à la définition donnée par Saussure lui-même et, à défaut, à celles rédigées par R. Godel (1957) et R. Engler (1968). La période d'attestation du mot-vedette, les sources dans lesquelles il figure, sa fréquence d'occurrence ainsi que les collocations dans lesquelles il apparaît sont également fournies. La classification ontologique du terme représente, il est clair, une donnée essentielle à laquelle s'ajoutent l'information concernant le domaine d'usage, le cas échéant le type d'événement dénoté, des traits sémantiques distinctifs ainsi qu'un vaste réseau de relations lexico-sémantiques. Le modèle lexical adopté accorde en effet une attention particulière aux liens existant entre les différents termes. Un ensemble de relations sémantiques — spécifiques du modèle *SIMPLE* et inspirées de la *structure des qualia* de la théorie du Lexique Génératif (Pustejovsky, 1995) — fournit une représentation structurée de la multi-dimensionnalité d'un sens. Ces relations offrent un cadre précis de la nature des relations (intra- et extra-catégorielles) qu'entretiennent les unités lexicales de la base de données. Elles explicitent leurs liens paradigmatiques : hyperonymiques, hyponymiques, méronymiques et holonymiques mais aussi, sur l'axe syntagmatique, les liens aux différents prédicats qui contribuent à éclairer leur sens ; ces derniers renseignent en particulier sur l'origine et la fonction de l'entité dénotée. L'information concernant les relations synonymiques, antonymiques et de dérivation morphologique est également présente. En outre, un ensemble de relations spécifiques a été créé, qui caractérise les liens particuliers existant entre les termes de ce domaine de la connaissance et, plus généralement, son organisation conceptuelle. Les termes prédicatifs figurant dans le lexique sont, de surcroît, dotés de leur structure argumentale avec indication du rôle sémantique et des restrictions sémantiques de leurs arguments.

Le lexique ainsi réalisé compte à ce jour 500 entrées lexicales, dont 370 noms, 118 adjectifs et 12 verbes. La base de données lexicales a été dans un premier temps gérée par le système de gestion MS Access. Plus récemment, dans un souci de standardisation et d'interopérabilité, une migration des données sur la plateforme Protégé-OWL a été effectuée. OWL est en effet le langage standard du W3C pour la représentation et le partage d'ontologies sur le Web. Le système de gestion Access n'a cependant pas été abandonné : le lexique réside à présent dans les deux plateformes, grâce à un logiciel assurant la parfaite concordance des informations saisies. Cette double résidence offre la possibilité d'exploiter de façon optimale les caractéristiques les plus remarquables des deux systèmes de gestion.

Une telle structuration informatisée des données lexicales offre de multiples avantages. En phase de création, elle permet un développement collaboratif rigoureux, même à distance, et le stockage d'une grande quantité de données ; elle favorise une représentation systématique 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 et 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, l'informatisation du lexique permet un accès simple et rapide aux données ainsi que la recherche ou l'inférence d'une vaste gamme d'informations. De plus, il est opportun de souligner que le développement et l'exploitation d'une telle ressource lexicale requièrent le simple apprentissage de l'outil de gestion et ne présupposent aucune connaissance informatique spécifique de la part de l'utilisateur.

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. Elles donnent la possibilité d'extraire des ensembles d'unités lexicales partageant une propriété ou une combinaison de propriétés, selon des critères établis par l'utilisateur pour les besoins de sa recherche. Ces requêtes étant effectuées dans un environnement formalisé qui garantit la complétude des données obtenues, leur résultat offre une vision plus approfondie et exhaustive de la nature componentielle et relationnelle du sens des mots.

À titre d'exemple, quelques une des requêtes nombreuses et variées pouvant être formulées sont exprimées ci-dessous en langage naturel :

- *Quels sont les termes définis comme résultant d'une association et quelle est leur classe conceptuelle d'appartenance ?*
- *Quelles sont les activités inconscientes du sujet parlant, ayant un résultat et qui sont du domaine de la synchronie ?*
- *Quels sont les termes caractérisés par le trait sémantique Mental et qui font partie du signe linguistique ?*
- *Quels sont les termes liés au terme 'signifiant', par quel type de relation, et à quelle classe conceptuelle appartiennent-ils ?*
- *Quels sont les termes utilisés par Saussure à partir de 1894 ?*

Enfin, les outils informatiques et les méthodologies adoptés pour le développement de ce thésaurus électronique sont entièrement portables puisque indépendants des langues étudiées et des champs d'application. Ils sont conçus afin d'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 Project<sup>ii</sup>, un thésaurus électronique de termes représentatifs extraits de la pseudo-Théologie d'Aristote est actuellement en phase préparatoire. Il sera fondé sur une version customisée de notre modèle lexical et sur l'expertise acquise lors du développement du lexique de la terminologie saussurienne.

## 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.
- 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., 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.
- Lenci, A., F. Busa, N. Ruimy, E. Gola, M. Monachini, N. Calzolari, A. Zampolli et al. (2000a). SIMPLE Linguistic Specifications LE4-8346 SIMPLE, Deliv. D2.1 & D2.2. Pisa : ILC and University of Pisa.
- Lenci, A. et al. (2000b). 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.
- Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA. : The MIT Press.
- 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.

<sup>i</sup> « Pour une édition numérique des manuscrits de Ferdinand de Saussure ».

<sup>ii</sup> ERC Ideas Advanced Grant 249431 *Greek into Arabic. Philosophical Concepts and Linguistic Bridges*.