Defining and Structuring Saussure's Terminology

Nilda Ruimy, Silvia Piccini & Emiliano Giovannetti

Keywords: *Saussure's terminology*, *computational lexicon*, *ontology*, *semantic relations*.

Abstract

In the framework of the Italian project '*For a digital edition of Ferdinand de Saussure's manuscripts*', an electronic thesaurus of Saussure's terminology is being built, which includes new terms extracted from recently found manuscripts. The lexical model on which it is grounded is a customized version of the SIMPLE model. In this paper, an overview of the customization process is provided, with a special focus on the steps taken for designing a domain-specific ontology as well as on the creation of additional semantic relations and features. Lexical entries are illustrated and the potential of a structured organization of semantic knowledge for gaining a wider understanding of the overall domain terminology is highlighted.

1. Introduction

Although Ferdinand de Saussure's thought and work have been widely diffused all over the world, the Genevan master never published his writing on general linguistic issues. As a matter of fact, his theses and theories were mainly reconstructed and interpreted by his colleagues and disciples Charles Bally and Albert Sechehaye (Saussure 1916), on the basis of student's notes. This is the reason why Saussure's own writings, published posthumously under the titles of *Recueil des publications scientifiques de Ferdinand de Saussure* and *Écrits de linguistique générale* are of paramount importance. They evidence the complexity of his philosophical and semiological system and the special attention Saussure paid to terminology, the lack of rigor of which he often deplored.

He himself used to change meaning to some concepts over time, to assign additional specific meanings to already existing terms, to use some expressions ephemerally, and even to forge new words. Saussure's terminology is therefore a very interesting and fruitful topic to explore with a view to getting a deeper understanding and hopefully to capturing potential developments of Saussurean thought.

The most relevant and comprehensive documentation in this regard is the *Lexique de la terminologie saussurienne*, written by Rudolf Engler and that dates back to 1968. Consequently, an updated study of Saussure's terminology including the *Écrits de linguistique générale* and taking advantage of modern information technology seems a most timely initiative.

To this aim, the CNR-ILC Research Unit is contributing to the Research Project of National Interest PRIN, coordinated by Prof. D. Gambarara and entitled '*For a digital edition of Ferdinand de Saussure's manuscripts*' with the creation of an innovative semantic 'thesaurus-lexicon' of Saussure's linguistic terminology. This compound appellation was adopted in view of the fact that the lexical knowledge base that is being built is much richer than a traditional thesaurus[1].

2. Selecting the starting model

Considering our objective, a first working phase was devoted to reviewing the available lexical models. Adapting an existing framework to domain-specific requirements was in fact deemed a more convenient option in terms of time consumption and cost, but above all in

terms of result quality. Today, well-founded lexical models designed during the last decade enable to build lexical databases with a multidimensional structuring of concepts and a large network of semantic links among terms. Grounded on models such as WordNet (Fellbaum 1998), EuroWordNet (Vossen 1999), ItalWordNet (Roventini et al. 2003), Framenet (Fillmore et al. 2003), Pattern Dictionary (Hanks and Pustejovsky 2005), SIMPLE (Lenci et al. 2000) and Brandeis Semantic Ontology (Pustejovsky et al. 2006), outstanding lexical resources for NLP were developed, which provide a rich description of word meaning with a view to retrieving and processing lexical data in texts.

To develop the very first electronic thesaurus of Saussure's terminology, we selected the SIMPLE conceptual model, with a view to customizing it into the domain-specific *SIMPLE_FdS* model. This model was in fact deemed the most adequate since, in the domain of Computational Lexicography, it distinguishes itself by some particularly valuable and innovative aspects. In the context of the eponym European project, this model - whose theoretical framework is the Generative Lexicon Theory (Pustejovsky 1995) - provided a semantic classification and a formal and structured representation of word meaning, as well as contextual semantic information, for twelve European languages. It thus turned to be a de facto standard and consequently strongly inspired the *Lexical Markup Framework* metamodel, which is the ISO standard for NLP lexicons (Francopoulo 2007).

SIMPLE has a flexible and modular architecture that enables to perform, at the desired granularity level, a rich and explicit semantic description that combines ontological classification and semantic valency (Ruimy et al. 2003). The genericity and systematicity features of the model and the coherent structuring of information answer the requirements for a fruitful exploitation of encoded data. In fact, SIMPLE allows an easy access to information and offers many interesting possibilities of retrieval, search and extraction of lexical data. Finally, this model, that was designed with a view to future expansions and specializations, displays properties of flexibility and versatility that lend it to being easily customized.

3. The conceptual level

The customization process started with the design of the core component of the *SIMPLE_FdS* model, that is a domain-specific lexical ontology[2] that would structure Saussure's terminology. Building an ontology amounts to modeling the conceptual schema of a domain knowledge according to ontology building principles. In the present case, the domain ontology was conceived in conformity with the architecture and design principles of the SIMPLE type system. It was therefore tailored to account for the different degrees of internal complexity of word meaning.

Combining bottom-up and top-down approaches, we started by imposing a rough, preliminary semantic classification on Engler (1968) and Godel's (1957) indexes of terms as well as on new representative terms extracted from the manuscripts under study. This permitted us to identify the main conceptual classes that were then defined and hierarchically arranged into a type system. Subsequently, they were generalized and/or specialized, trying to keep a balance between too general classes and too fine-grained distinctions. The internal structure of the types, that is their attributes and values, was then defined. Following our guiding model, types were organized along hierarchical and non-hierarchical conceptual relations and distinguished into one-dimensional and multidimensional. In the latter types, the multidimensionality of meaning is captured by means of the four roles of the *Qualia Structure* (*formal, constitutive, agentive* and *telic*). In the ontology, *qualia* roles define the distinctive properties of semantic types and differentiate their internal semantic constituency.

The current version of the type system may still undergo some adjustments that will clearly be performed in full compliance with the ontology building principles of the archetypal model.

The *SIMPLE_FdS* ontology has been ported in OWL (Web Ontology Language). The advantage of having an OWL version of the ontology is threefold: i) since OWL has model theoretic formal semantics, it is possible to run a number of ontology and consistency tests (duplicated restrictions, multiple asserted parents, etc.); ii) a variety of visualization plugins for the OWL editor Protégé is available and allows to simplify both the editing of the ontology and its consultation; iii) OWL is the standard language to represent and share an ontology on the Web.

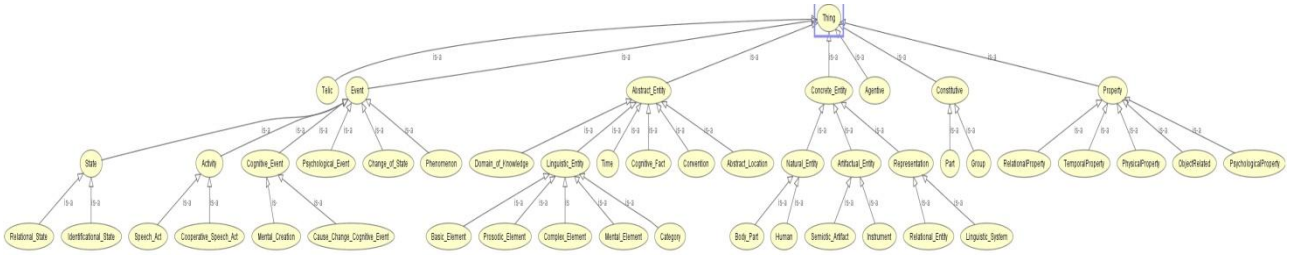

Figure 1. *Simple_FdS* ontology, version 0.2[3]

Besides the ontology, two other building blocks of the *SIMPLE_FdS* model, that is semantic features and semantic relations, enable to express a wide typology of information characterizing a word's semantic content.

Semantic features express information types such as the domain of use of a word, distinctive properties that cut across the type hierarchy and thus allow to cluster word senses whatever their ontological classification, and traits interpreting meaning dimensions clearly perceived in a word semantic content but hardly expressible within a semantic relation.

Semantic relations, the heart of the lexical model, are expressed as triplets: <*source semantic unit, relation, target semantic unit*>. Their bulk and core is taken from SIMPLE *Extended Qualia Structure*, which is the outcome of a revisitation of *Qualia Structure*. *Extended Qualia* relations allow to express the orthogonal meaning dimensions coexisting in the lexical semantics of a word sense and to structure its relationships to other lexical units, on both paradigmatic and syntagmatic axes. They provide a fine-grained knowledge on the identity, constituency, origin and function of an entity.

Other relationships are also represented, for example synonymy, antonymy, morphological derivation, as well as relationships holding between events and participants and among co-participants in events.

The *SIMPLE_FdS* model encompasses those SIMPLE original features and relations deemed suitable for our domain of interest as well as new specific features and relations created to account for term properties and relationships peculiar to the conceptual organization of the domain knowledge.

Some sixteen new features encode domain-specific subject fields, information on source, frequency and period of attestation of a term.

So far, sixteen specific relations were created, which are illustrated in table 1.

Table 1. Examples of domain-specific semantic relations.

signifié **hasPreviousDenomination** *concept*	*temps homogène* **hasSubsequentDenomination** *moment*
parasème **hasNearSynonym** *terme*	*morphologie* **hasOtherDenomination** *théorie des signes*
syllabe **pertainsTo** *phonologie*	*chaînon* **belongsTo** *synchronie*
arbitraire **duality** *motivation_relative*	*formatif* **isOpposedTo** *radical*

sôme **directlyImplies** *antisôme*	*concept verbal* **indirectlyImplies** *image auditive*
préfixe **precedes** *radical*	*désinence* **follows** *thème*
chaînon **isEdSubstitutedWith** *anneau*	*anneau* **substitutesFor** *chaînon*
paraplasme **hasEndonym**[4] *métaplasme*	*trésor* **abstractLocation** *pensée*

4. The terminological level

The descriptive means illustrated enable to create rich lexical entries (cf. fig. 2) in which each single sense of a (one- or multi-word) lexeme is encoded. A word sense (or *semantic unit*) is associated to a large set of structured and fine-grained formalized information concerning a wide range of semantic aspects. Special emphasis is given to the relationships holding among terms.

Privileging the terms extracted from Saussure's authentic writings, we created up to now 375 lexical entries (284 nouns or noun phrases, 70 adjectives and 21 verbs)[5].

id = "UsemD8388parasème"
Naming = "parasème"
Saussure definition = "Pour un mot quelconque faisant partie de la langue un second mot, même n'ayant avec le premier aucune parenté, est un parasème. La seule et simple qualité de parasème est de faire partie d'un même système psychologique de signes. " 3313.2
Definition = "Signe (voir sème) considéré dans ses rapports avec le système ; voir terme. " (Engler, 38).
Semfeaturelist=" SemanticType Relational_Entity SuperType Representation Domain Linguistics Domain Semiology Plus_ Conventional Plus_Semiotic IndirectTelicYes"
Relations:
"isa" "USem5signe" (Relational_Entity)
"isaPartOf" "UsemD4033système" (Constitutive)
"directlyImplies" "UsemD17valeur" (Relational_State)
"hasEndonym" "UsemD36sème" (Relational_Entity)
"hasNearSynonym" "UsemD60001terme" (Relational_Entity)"

id = "USemD24signifiant"
Naming = "signifiant"
Definition = "Le signifiant et le signifié sont les deux éléments composant le signe. Le signifiant est auditif, le signifié conceptuel." (Godel: Cahiers D, 211).
Collocations = "signifiant linguistique"
Semfeaturelist "SemanticType Concrete_Entity SuperType Thing Domain Linguistics Plus_ Conventional Plus_Semiotic Plus _Sound IndirectTelicYes "
Relations:
"isa" "USemD1élément" (Constitutive)
"isaPartOf" "USem5signe" (Relational_Entity)
"isaPartOf" "USemD27signe_linguistique" (Relational_Entity)
"concerns" "USemD37audition" (Psychological_Event)
"directlyImplies" "USemD35signifié" (Mental_Element)
"indirectlyImplies" "USem35pensée" (Mental_Element)
"indirectlyImplies" "USemD23sens" (Mental_Element)
"hasProperty" "USemD8376linéarité" (Property)
"hasSynonym" "UsemD9139sôme" (Concrete_Entity)
"hasPreviousDenomination" "USemD88image_acoustique" (Concrete_Entity)

Figure 2. Examples of lexical entries.

The whole set of entries is stored in a database and managed through a system that allows a concurrent and easy access to data for creating, storing, consulting and modifying lexical entries. The database management system includes a set of SQL[6] queries that enable to perform quality checks on the formal consistency of data. Most importantly, it supports advanced querying functionalities for a quick, efficient and comprehensive retrieval of lexical information.

Table 2. Query for a pair: specific semantic relation - specific target term.

Source semantic unit	Naming	Relation	Target semantic unit
USem42agglutination	agglutination	**belongsTo**	**USem027diachronie**
USem441étymologie	étymologie	**belongsTo**	**USem027diachronie**
USemD9021métaplasme	métaplasme	**belongsTo**	**USem027diachronie**
USem4419mutabilité	mutabilité	**belongsTo**	**USem027diachronie**
USem15phonétique	phonétique	**belongsTo**	**USem027diachronie**
USemD9057reconstruction	reconstruction	**belongsTo**	**USem027diachronie**
…	**…**	**belongsTo**	**USem027diachronie**

Information retrieval queries may be formulated using any single piece of encoded information, be it a semantic relation, a feature or a lexical unit. For example, a query for a

given semantic relation (either coupled with a specific target term or not) renders all pairs of terms linked through this relation. Through a query for a lexical unit, all terms connected to it are retrieved and classified according to the type of relationship they hold with the keyword. Querying for a semantic feature, all involved terms are captured, whatever their ontological classification. Besides, coherence and completeness of results are ensured by the formalized framework in which queries are performed.

5. Concluding remarks

Such an electronic thesaurus-lexicon, based on a multidimensional structuring of concepts and a large network of semantic relations among terms, is, in our opinion, a most valuable lexical research tool. It offers many interesting possibilities of investigating lexical data. In the lexicon, the overall structure of Saussure's terminology is made explicit and the semantic import of its component terms as well as the nature and relevance of their relationships are defined. The structured organization of lexical information, which highlights the componential and relational nature of word meaning, should provide a deeper knowledge of the overall domain terminology and might therefore contribute to a better understanding of the author's original thought, and to shade light on some of his complex reflections.

The tools and methodologies developed for building this electronic thesaurus are fully portable and are intended to be used in similar domains. Actually, in the framework of an ERC Advanced project, an electronic semantic thesaurus of Arabic terms extracted from the pseudo-*Theology of Aristotle* is presently being built, which is grounded on a specifically customized version of the root model.

Notes

[1] A thesaurus is generally defined as an ordered collection of terms related by hierarchical, associative and synonymic relations. These terms make up the lexicon of a domain knowledge and are used for text indexing and information retrieval within that domain. Thesauri may be based on a unidimensional or multidimensional classification system.

[2] We adopt the definition of the term 'ontology' provided by T.R. Gruber (1993: 200): 'An ontology is an explicit specification of a conceptualization.'

[3] For a higher quality view of the ontology, click on the following link http://tinyurl.com/72mhcnv.

[4] This semantic relation was inspired from (Cruse, 1986 : 123).

[5] The first release of the semantic thesaurus-lexicon of Saussure's terminology is due for the end of 2012.

[6] **S**tructured **Q**uery **L**anguage, a programming language designed for managing data in relational databases.

References

Cruse, D. A. 1986. *Lexical Semantics*. Cambridge: Cambridge University Press.

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.

Godel, R. 1957. *Les Sources Manuscrites du Cours de Linguistique Générale de Ferdinand de Saussure*. Genève: Droz.

Engler, R. 1968. *Lexique de la Terminologie Saussurienne*. Comité international permanent des linguistes. Publication de la commission de terminologie. Utrecht-Anvers: Spectrum.

Gruber, T. R. 1993. 'A Translation Approach to Portable Ontologies.' *Knowledge Acquisition* 5.2: 199–220.

Hanks, P. and J. Pustejovsky 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. ILC and University of Pisa. Pisa.

Pustejovsky, J. 1995. *The Generative Lexicon*. Cambridge MA: The MIT Press.

Pustejovsky, J. et al. 2006. 'Towards a Generative Lexical Resource: The Brandeis Semantic Ontology.' In *Proceedings of LREC 2006*. Genova, 1702–1705.

Roventini A. et al. 2003. 'ItalWordNet: Building a Large Semantic Database for the Automatic Treatment of Italian.' In A. Zampolli, N. Calzolari, L. Cignoni (eds.), *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 A. Zampolli, N. Calzolari, L. Cignoni (eds.), *Computational Linguistics in Pisa, Special Issue*. Vol.XVIII-XIX. Tomo II. Pisa-Roma: IEPI, 821–864.

Saussure, F. de 1916. *Cours de Linguistique Générale*. Ch. Bally and A. Sechehaye (eds.). Lausanne-Paris: Payot.

Saussure, F. de 1922. *Recueil des publications scientifiques de Ferdinand de Saussure,* Ch. Bally and L. Gautier (eds.). Lausanne: Payot.

Saussure, F. de 2002. *Écrits de linguistique générale.* S. Bouquet and R. Engler (eds.). Paris: Gallimard.

Vossen, P. (ed.) 1999. *Eurowordnet. A Multilingual Database with Lexical Semantic Networks*. Dordrecht: Kluwer.