Personalized Semantic Resources
The SemComp Project Presentation and Preliminary Works

Alexandre Labadié[1], Stéphane Ferrari[1] and Thibault Roy[2]
[1]*GREYC, Caen, France*
[2]*Noopsis, Caen, France*

Keywords: Personalized Semantic Resources, Componential Semantic, RDF.

Abstract: This paper presents the computational aspects of the SemComp project, a multidisciplinary collaboration aiming at observing how interacting with documents acts on knowledge acquisition. It is based on a model for personalized semantic resources inspired from componential linguistics. The paper describes the advances in both the computational model's definition as well as its implementation in a Web oriented application. Functionalities and technical choices are presented with regards to the expected experiments.

1 INTRODUCTION

In this paper, we present the SemComp project, an ongoing research project funded by the French Region *Basse Normandie*. It aims to experiment a model for semantic representation in the applicative context of enhancing personal access to documents. The experimentation will be realized through a Web-oriented application, in teaching environments or for cultural tourism purpose. SemComp is a multidisciplinary collaborative project involving linguists, psychologists and computer scientists. We thus intend to reach multiple goals: testing a new implementation of a linguistic model where personal interpretation is central, collecting data to observe how interacting with documents acts on lexical and semantic knowledge acquisition, testing a method for personalized access to information.

In section 2, we briefly present the model for semantic representation. It is a simplification of a linguistic approach to componential semantics for applicative purposes: casual Web users with or without linguistic knowledge may use it. In section 3, we detail some aspects of the implementation of Personalized Semantic Resources (PSR in the following) using Web semantic tools and standards. To conclude, we present future works and evolution of the project.

2 SEMANTIC MODEL

SemComp stands for *Sémantique Componentielle*, "Componential Semantics".

2.1 Motivation

The model for semantic representation and, as a consequence, for semantic analysis, is inspired by the structural linguistic approach. It relies on the main notions of the "Componential Semantics" and the "Interpretative Semantics" as proposed in, e.g., (Greimas, 1966; Pottier, 1992; Rastier, 1987). These notions are detailed in section 2.2.

In the last decade, NLP works based on a similar approach have already been realized. In (Beust et al., 2003), a componential model has been tested for semantic analysis, and more specifically for metaphors detection and interpretation in domain-specific corpora. In (Roy and Ferrari, 2008), the same model as well as graphical tools are used to provide a user with a personal access to textual information. The applications were in both cases designed for experts rather than casual Web users. Different other works, like (Valette and Slodzian, 2008; Kanellos and Mauceri, 2008), also proposed the use of "Interpretative Semantics" for information retrieval application.

In the SemComp project, we propose a modified, simplified version of this model for its use by casual Web users. Relaxing some constraints in the structure, we consider there are strong similarities with some Web 2.0 tools such as folksonomies, tags and social bookmarking.

Labadié A., Ferrari S. and Roy T..
Personalized Semantic Resources - The SemComp Project Presentation and Preliminary Works.
DOI: 10.5220/0004539501640169
In *Proceedings of the International Conference on Knowledge Engineering and Ontology Development* (KEOD-2013), pages 164-169
ISBN: 978-989-8565-81-5

2.2 Linguistic Model

***Sèmes*.** The central notion of the "Componential Semantics" approach is to describe lexical units with semantic features. These features, called *"sèmes"*, are theoretically supposed to describe the possible interpretations of a lexical unit.

Some *sèmes* are generic ones, representing parts of meaning shared by the lexical unit and the ones with close meaning. For instance, *chair* and *sofa* will share *sèmes* because they both can mean "a sort of seat". One may consider *sèmes* like */physical object/, /crafted object/, /piece of furniture/, /seat/* to explain these meanings.

Some *sèmes* are specific ones, used to distinguish between close meanings. *Chair* could be differentiated from *sofa* using a *sème* such as */seat without arms/*, while *chair* and *sofa* could themselves be differentiated from *stool* using a *sème* such as */seat without a back/*.

Previous works such as in (Beust et al., 2003; Roy and Ferrari, 2008) proposed to represent the specific *sèmes* using attributes and values to code their differential role: */seat's back: yes/* for *chair* and *sofa*, */seat's back: no/* for *stool*; */seat's arms: yes/* for *sofa*, */seat's arms: no/* for *chair* and *stool*. These constraints are strong ones, requiring a high level of expertise for describing a whole domain.

***Sèmes* in SemComp: Free Semantic Features.** In the SemComp project, we propose to let the user describe the semantic features freely. We make the hypothesis casual users will mostly describe generic *sèmes* in order to retrieve documents related to their hobbies and interests. The application in view (see 2.3) will allow a user to build queries using *sèmes* rather than lexical (or graphical) units only. For casual users, we make the assumption *sèmes* will be used as tags for text classification: rather than tagging texts, users will tag words themselves.

Though, it will still be possible for an expert user to propose sets of differential *sèmes* if necessary. In the application, this will appear only as an advanced functionality. With the previous examples, a user will be able to build a differential set including the *sèmes /seat without a back/* and */seat with a back/*, in order to enhance the results of a query using the *sèmes /seat/* and */seat without a back/*: the application will automatically consider the *sème /seat with a back/* as irrelevant, though describing units with close meanings.

Isotopies. The second central notion of the "Interpretative Semantics" is the one of "isotopy" for contextual interpretation. An isotopy is the redundancy of a *sème* in a textual zone. It is closely related to the notion of topic. When using numerous *sèmes* to describe a domain, as for the experiment on metaphors in (Beust et al., 2003), it has been proved isotopies help activating or deactivating *sèmes*: in the context of economic news, meteorological terms such as *thermometer* and *barometer* can be interpreted as measuring or prevision tools for stock markets, deactivating *sèmes* specific to the meteorology (the units they use, the phenomenon they measure, etc.). French linguists propose the terms of *actualisation* (the action of activating a *sème* in a specific context) and *virtualisation* (the action of deactivating a *sème* in a specific context) to describe these interpretation processes.

2.3 Application in View

Based on this simplified linguistic model, we plan to develop an application allowing users to create, manipulate and use their personal points of view on different domains for consulting documents. Different psychological and NLP experimentations are expected in the SemComp project. The use of Personalized Semantic Resources (PSR in the following) will be tested in the following applicative contexts : (1) students consulting teachers on-line courses ; (2) students consulting the Web for a class project ; (3) tourists looking for cultural activities in Normandy.

Experiments (1) and (2) are scheduled in a short term (1yr), while (3), requiring inclusion of other NLP tools, is scheduled in a longer term (2yrs). In (1) and (2), we expect to observe how students acquiring new knowledge on a domain modify their PSR. In (3), we intend to experiment on casual users, and include sharing of PSR to test if this model can lead to real Web applications.

In (1), the collection of documents is closed, limited to the documents provided by a teacher in the scope of a course. In (2), the collection must first be retrieved from the Web, using a search engine, which require to translate the user request from *sèmes* to written forms. Next section presents the first developments centered on the PSR, as well as more details on the application functionalities which will be used in the first experiments (1) and (2).

3 PERSONALIZED SEMANTIC RESOURCES (PSR)

Based on the model previously presented, we are currently developing a set of resources and an application linked to it aiming to achieve three goals:

1 to propose an exhaustive, yet flexible, representation for PSR;

2 to allow model implementation in an applicative context ;

3 to track how users build their semantic resources, for experimentation purposes.

3.1 UML Model

The simplified class diagram (figure 1) can easily be divided into three parts corresponding to our three goals : the semantic resources (PSR), their connexions with documents (applicative context) and the interactions (tracked for experimentations).

The center of the diagram represents the semantic resources a user can build. The heart of the model is the lexical entry which is linked to multiple written forms (at least one) which compose it. One of the written forms is the lexical entry representative. A lexical entry can be linked to one or more features groups, each one representing a different meaning. These groups are composed of semantic features (*Sèmes*). The semantic features are not limited to textual representations and can be images, sounds, *etc.* Features groups can be of different types, the most basic one is a meaning of a lexical entry, but the model allows to create other types of group to identify specific properties of some semantic features (for instance a "differential set of *sèmes*" as illustrated in the previous section). Following the same idea features groups can be linked by pairs to represent lexical or semantic relations (hyperonymy, synonymy, *etc.*). Lastly, lexical entries, semantic features and features groups are linked to a viewpoint of a specific user on a domain.

The top (and right-top) of the class diagram is devoted to link the PSR model to documents. A query can be composed by selecting different semantic features. Features that can be activated or deactivated by the user when she meets an occurrence of a written form in a document. The main purpose here is to propose a model that allows a user to create queries using her own personalized semantic representation of a domain (using semantic features to create and expand queries). But our model also allows to reorder the returned collection of documents. By activating and deactivating semantic features linked to words occurrences in the documents, the user will refine her search.

Lastly, the bottom of the class diagram allows us to track each interaction the user has with her PSR. This will be used in psycho-linguistic experiments to uncover the building process of a semantic representation of a domain and its uses. Of course, the tracking process will only take place during experiments and users will be aware of it.

3.2 Application Expected Functionalities

For Users. In addition to a user friendly manipulation of our model, our application will allow the user to use her own PSR to improve her Web or closed collections of documents research. The user will be presented with the semantic features of her own PSR and will compose queries with them. By expanding these "*Sème* formulated queries", the application should retrieve documents closer to the user point of view on a domain than a more classic approach. We also intend to allow users to share parts of their PSR and to expand their owns with parts shared by other users.

For Experimenters. As described in section 3 our model allows to track the user interactions while she is building and using her PSR. The final application should show us the whole process of building a PSR by the user, in interaction with document browsing. It should also allow us to navigate between every step of this process: when tracking is active, the application keeps everything in the RDF repository (see following section for technical choices), archiving any modified or deleted instance. We also intend to use graph similarity measures to compare different users, domains, *etc*.

3.3 Technological Choices, Implementation and Preliminary Results

We chose to use Web semantic tools and standards to implement our PSR. We turned to RDF[1], which graph-like representation is closer to our model than classical relational ones. Figure 2 is an example of some semantic data organized as a multi-user PSR. It show us two users sharing one domain, with common and distinct semantic features, lexical entries and such (these are simplified data for test and example purposes). Purple triangles are classes from our model and green labels are their implementation. In figure 2, the aeronautics domain is linked to two points of view from two different users (*Alexandre* and *Stéphane*). Each point of view has lexical entries, semantic features (aka. *sèmes*), written forms, *etc.*. For instance, *Boeing* is the written form *graphie_00002* representing the lexical entry *lexie_00002*. One of its meanings (a plane) is described by the features group

[1]W3C RDF reference site: http://www.w3.org/RDF/

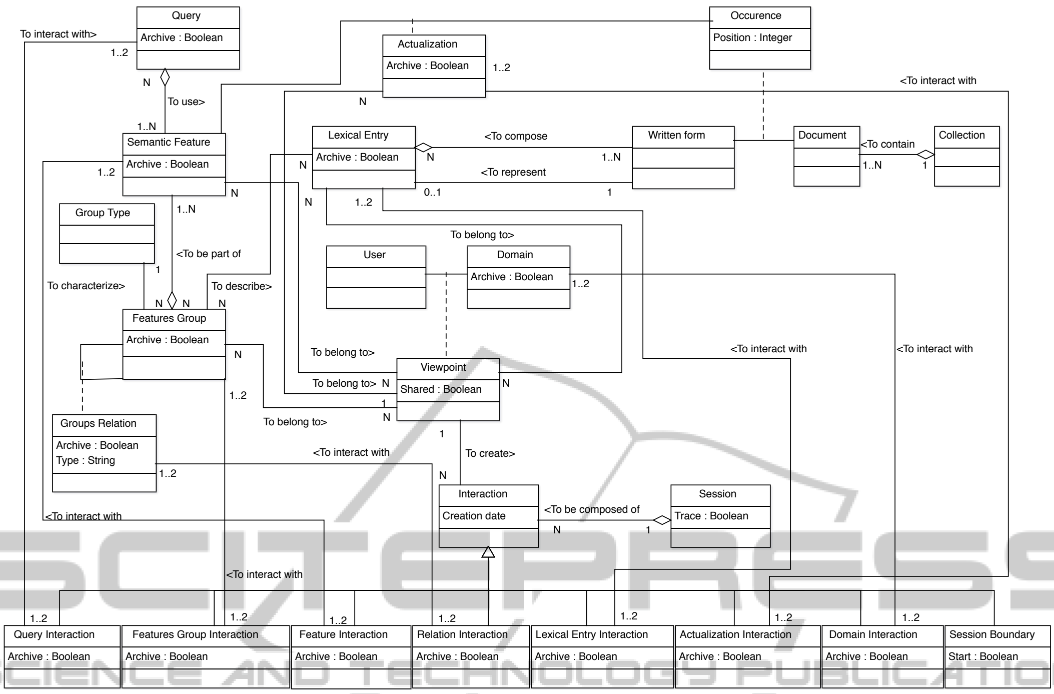

Figure 1: PSR UML class diagram.

groupe_semes_00004 entitled */L'avion Boeing/* (the Boeing plane, it is a common metonymy in French to use the name of the company for one of its planes). This features group contains the two *sèmes /vole/* (can fly) and */transporte/* (can carry). The meaning for Boeing as a company is not described in this graph.

Some of these data are common to both points of view, others are specific to one only. These test data are the advanced form of the PSR, where users can share parts of their PSR. The first set of experiments[2] will not allow to share their PSR, and will only allow them to access their personal resources and some common resources extracted from a dictionary. Resource sharing will come in a second experimental phase [3].

The prototype is currently implemented both as a JAVA Web service and a front end Web application. It is based on the open source RDF base OpenRDF Sesame[4]. In due time, it should be integrated in a bigger Web framework including other NLP applications interacting with each other. First, we intend to propose a simple Web interface to experiment directly on the model. In a second time, as other NLP applications will be integrated into our framework, we should develop a more complete Web gateway offering extended services.

At the time we are writing these lines, the model has been implemented and deployed in an OpenRDF Sesame repository, the JAVA Web service and the Web interface are currently being developed. The application prototype can be tested here: https://semcomp.info.unicaen.fr/.

4 FUTURE WORKS

In this paper, we presented some aspects of the ongoing SemComp project. It aims at observing how users build their lexical and semantic knowledge while interacting with documents. For this purpose, we proposed a simplified model for Personalized Semantic Resources based on the linguistic approach to Componential Semantics. The main idea is to associate lexical entries with features called *sèmes* to represent parts of their meaning or interpretation. We presented the current implementation of this Personalized Semantic Resources in a Web-oriented application using one of the latest Web semantic tools at our disposal (RDF triple stores). We are currently developing the Web client interface to provide users with two main functionalities : defining their PSR, building requests using *sèmes* to search information through doc-

[2]Scheduled in September 2013 in Caen University, France, for online courses access and in Louis Liard high school, Falaise, France, for Web access.

[3]Scheduled in March 2014 for cultural tourism in the *Basse Normandie* Region

[4]OpenRDF Web site: http://www.openrdf.org/

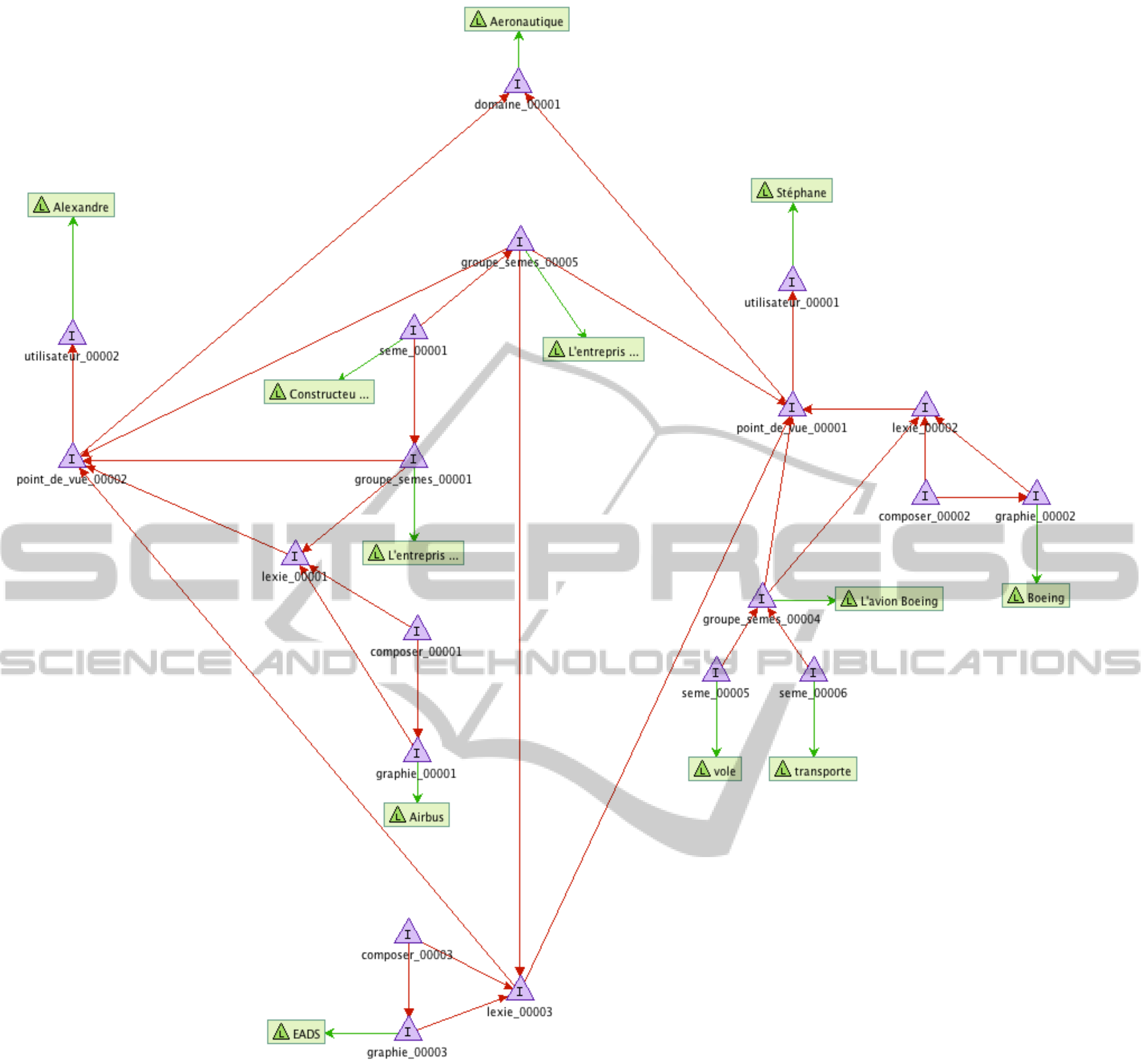

Figure 2: Part of the graph / PSR representation of the aeronautics domain for a pair of users.

uments.

In order to allow casual Web users to define their PSR, the underlying linguistic notions are hidden in this interface : the user is just asked to define her own tags (the *sèmes*) and tag words she finds relevant for her task the way she would tag relevant documents in a Web2.0 application. The user can also build request using *sèmes* only to search a specific information. A back-office module is dedicated to the translation of such requests into lists of written forms to query a search engine. A second back-office module validates and sorts the retrieved documents with regards to the user's PSR and the initial *sèmes* request. We expect users will refine their PSR while discovering information relevant for their task in the retrieved documents, e.g. new lexical entries or new meaning features. In the psycho-linguistic experiments, the interactions with both the PSR and the documents will be tracked in order to observe the knowledge acquisition process.

The first two experiments will involve students searching the Web for a class project or accessing their teacher's online courses. In both cases, the users are expected to acquire knowledge. A long-term experiment is also expected. PSR would be integrated in a larger "cultural tourism" NLP application for Web users. It will help the user to find which cultural events are happening during her stay in a specific location. This latter experiment aims to test the model in a real application context, and not only for psycho-linguistic experiment. We intend to use the PSR to help the user describing her interests and to enhance the matching between the found events and her interests.

REFERENCES

Beust, P., Ferrari, S., and Perlerin, V. (2003). Nlp model and tools for detecting and interpreting metaphors in domain-specific corpora. In Archer, D., Rayson, P., Wilson, A., and McEnery, T., editors, *Proceedings of the Corpus Linguistics 2003 conference*, volume 16 of *UCREL technical papers*, pages 114–123, Lancaster, U.K.

Greimas, A. J. (1966). *Sémantique structurale : recherche et méthode*. Larousse.

Kanellos, I. and Mauceri, C. (2008). Une conscience interprétative face à un univers de textes. arguments en faveur d'une analyse de données interprétative. *Syntaxe & sémantique*, (9). Textes, documents numériques, corpus. Pour une science des textes instrumentée. Etudes publiées sous la direction de Mathieu Valette.

Pottier, B. (1992). *Sémantique générale*. Presses Universitaires de France.

Rastier, F. (1987). *Sémantique interprétative*. Presses Universitaires de France.

Roy, T. and Ferrari, S. (2008). User preferences for access to textual information: Model, tools and experiments. In Wallace, M., Angelides, M., and Mylonas, P., editors, *Advances in Semantic Media Adaptation and Personalization*, pages 285–306. Springer.

Valette, M. and Slodzian, M. (2008). Sémantique des textes et recherche d'information. *Revue française de linguistique appliquée*, XIII(1):119–133.