

Moving to dynamic computational lexicons with LeXFlow

Claudia Soria[◦], Maurizio Tesconi⁺, Francesca Bertagna[◦], Nicoletta Calzolari[◦],
Andrea Marchetti⁺, Monica Monachini[◦]

[◦]CNR - Istituto di Linguistica Computazionale

⁺CNR – Istituto di Informatica e Telematica

Area della Ricerca di Pisa, via Moruzzi 1

56124 Pisa, Italy

{claudia.soria,francesca.bertagna,nicoletta.calzolari,monica.monachini}@ilc.cnr.it

{maurizio.tesconi,andrea.marchetti}@iit.cnr.it

Abstract

In this paper we present LeXFlow, a web application framework where lexicons already expressed in standardised format semi-automatically interact by reciprocally enriching themselves. LeXFlow is intended for, on the one hand, paving the way to the development of dynamic multi-source lexicons; and on the other, for fostering the adoption of standards. Borrowing from techniques used in the domain of document workflows, we model the activity of lexicon management as a particular case of workflow instance, where lexical entries move across agents and become dynamically updated. To this end, we have designed a lexical flow (LF) corresponding to the scenario where an entry of a lexicon A becomes enriched via basically two steps. First, by virtue of being mapped onto a corresponding entry belonging to a lexicon B , the entry_(LA) inherits the semantic relations available in lexicon B . Second, by resorting to an automatic application that acquires information about semantic relations from corpora, the relations acquired are integrated into the entry and proposed to the human encoder. As a result of the lexical flow, in addition, for each starting lexical entry_(LA) mapped onto a corresponding entry_(LB) the flow produces a new entry representing the merging of the original two.

1. Introduction

This paper presents LeXFlow, a framework for the semi-automatic management of lexical entries. As it will become clearer through the paper, LeXFlow is intended to provide the language resource community with an architectural and practical framework enabling dynamic, semi-automatic integration of language resources, exemplifying the particular case of semantic computational lexicons. In a broader sense, the long-term goal of our project is to be intended as a step towards language resource interoperability. Since years a new generation of language resources has been called for, where content is dynamically augmented by resorting to heterogeneous sources (Calzolari & Soria, 2005). In order to attain better coverage, it should be possible for language resources to be automatically maintained and augmented, possibly by resorting to other sources than human, introspective knowledge and integrating the knowledge either already explicitly encoded in other resources or implicitly conveyed by corpora.

Computational lexicons aim at providing an explicit representation of word meaning, so that it can be directly accessed and used by computational agents. In the last decade, many activities at European level and worldwide have contributed to substantially advance knowledge and capability of how to represent, create, maintain, acquire, access, and share large lexical repositories. However, most existing lexical resources do not have enough coverage, not only for practical reasons, but also for more structural and inherent reasons. No individual “static” resource can ever be adequate and satisfying, neither in extension (since it cannot cover new formations, or all the possible domains) nor in depth (since it cannot provide all the necessary and useful linguistic information, not even for the existing lexical entries).

The computational lexicon community since many years is thus increasingly calling for a change in perspective on computational lexicons: from static

resources towards *dynamic multi-source entities*, integrating and harmonizing the linguistic information coming from different sources, where lexical content is co-determined by automatically acquired linguistic information from text corpora and from the web. A different scenario is thus envisaged, where acquisition tools are able to increase the repository with new words/terms, possibly their definitions, domain, etc., from digital material, to learn concepts from text, and to tailor resources to specific needs.

2. Background

The realization of such a challenging endeavour obviously requires radically new approaches at various levels, among which standardisation plays a central role. Important and extensive efforts have been and are being made towards the extension and integration of existing and emerging open lexical and terminological standards and best practices. Among them, we mention EAGLES (Sanfilippo *et al.* 1999), ISLE (Calzolari, Lenci, & Zampolli, 2001), TEI, OLIF, Martif (ISO 12200), Data Categories (ISO 12620), ISO/TC37/SC4, LIRICS (Francopoulo *et al.*, in press). An important achievement in this respect is the MILE, a meta-entry for the encoding of multilingual lexical information (Calzolari *et al.*, 2003). On the other side, very little has been made towards the development of new methods and techniques for enabling the dynamic paradigm of lexical resources. Some initial steps are also made to realize frameworks enabling inter-lexica access, search, integration and operability. An example is the Lexus tool (Wittenburg & Kemps-Snijders, 2006), based on the Lexical Markup Framework (Romary *et al.*, in press), that goes in the direction of managing the exchange of data among large scale lexical resources. A similar tool, but more addressed to the collaborative creation of lexicons for endangered language, is SHAWEL (Gulrajani, 2002). But the general impression is that little has been made towards the development of new methods and techniques for the concrete interoperability

among lexical resources. The aim of LeXFlow is to fill in this gap.

3. LeXFlow

We believe that an essential step towards the realization of the dynamic paradigm of lexical resources is closely related to the development of an appropriate framework for computational lexicons where lexical entries behave as semi-independent entities, that dynamically modify and update their content on the basis of the integration of knowledge coming from different sources, where the sources can be indifferently represented by human agents, other lexical resources, or applications for the automatic extraction of lexical information from texts. This scenario has at least two strictly related prerequisites: on the one hand, it assumes that existing lexicons can be mapped to a standard form enabling the overcoming of their respective differences and idiosyncrasies, thus making their mutual comprehensibility a reality. On the other, it calls for the provision of an architectural framework for the effective and practical management of lexicons, by providing the communicative channel through which lexicons can really communicate and share the information encoded therein.

3.1. General architecture: the metaphor of “lexical workflow”

A similar approach is adopted in the domain of document workflow (DW). In document workflow systems, all activities made by some agents result in document compilation. It can be viewed as the automation and administration of particular documents procedures (Marchetti *et al.*, 2001). In other words, a DW can be seen as a process of cooperative authoring where the document can be the goal of the process or just a side effect of the cooperation. Through a DW, a document life-cycle is tracked and supervised, continually providing document compilation actions control.

Similarly, the management of computational lexicons can be described as a flow of lexical entries. A lexical entry is modeled as a document moving through different agents, with clear-cut roles, acting over different portions of each entry. Following this metaphor, LeXFlow is conceived as a metaphoric extension and adaptation to computational lexicons of XFlow, a framework for the collaborative management of document workflows (Marchetti, Tesconi, & Minutoli, 2005).

In this environment there are two types of agents: *internal agents* are software actors providing general-purpose activities useful for any workflow and hence are implemented directly into the system, while *external agents* are human or software actors that perform activities dependent from a particular lexical workflow (LW). Internal agents perform general functionalities such as creating/converting an entry belonging to a particular LW, populating it with some initial data, duplicating an entry to be sent to multiple agents, splitting an entry and sending portions of information to different agents, merging duplicated entries coming from multiple agents, aggregating fragments, and finally terminating operations over the entry. External agents basically execute some processing using the already available content of the entry and populate it with lexical information. In our demonstrative LW, a particular type of external agent is

represented by an application that acquires information about *part-of* relations by identifying syntactic constructions that are often used to express such relations. Other external agents are one or more compilers and one or more responsables for quality control, who basically check the output of the previous agent(s), validate it, and send the document to the next agent(s) (see Fig. 2 below). In order to account for the peculiarities of lexicon encoding and management, XFlow has been extended and specialized.

In the LeXFlow framework the workflow of lexical entries is described by a new XML application called XFlowML (XFlow Markup Language), largely based on XSLT Processing Model. XFlowML describes a workflow using an agent-based approach. Each human or software agent can participate to the workflow with one or more roles, defined as XPath expressions, based on a hierarchical role chart. An XFlowML document contains as many templates as are the agent roles participating in the workflow. The selection of the templates will establish the order with which the agents will receive the lexical entry. The document workflow engine constitutes the runtime execution support for the document processing by implementing the XFlowML constructs. To this end, at first we have defined the logical schema of a lexical entry and the contextual domain of the document workflow including all human and software agents cooperating, with different roles, to the compilation of lexical entries. Finally we have formalized the procedural rules and the access control rules (XFlowML) of lexical entry compilation.

A prototype of LeXFlow has been implemented with an extensive use of XML technologies (XML Schema, XSLT, XPath, XForms, SVG) and open-source tools (Cocoon, Tomcat, MySQL). It is a web-based application where human agents interact with the system through an XForms browser that displays the document to process as a web form whereas software agents interact with the system via web services.

3.2. Representing lexical entries: the MILE lexical model

In order to ensure interoperability, an essential prerequisite is the requirement that lexicon entries be encoded in a shared, standard format. We have chosen to use the MILE (Multilingual ISLE Lexical Entry, Calzolari *et al.* 2003) as a standardized model to describe the entries belonging to different lexicons. The MILE is a general architecture devised for the encoding of multilingual lexical information, a meta-entry acting as a common representational layer for multilingual lexicons, by allowing integration and interoperability between different monolingual lexicons. Although primarily devised for multilingual lexicons, the MILE can also be applied to mono-lingual lexicons. MILE-conformant lexical entries can be built by lexicon and application developers by means of the overall MILE Lexical Model (MLM). According to the model, the monolingual component on the vertical dimension is organized over three different representational layers which allow to describe different dimensions of lexical entries, namely the morphological, syntactic and semantic layers. Moreover, an intermediate module allows to define mechanisms of linkage and mapping between the syntactic and semantic layers.

Within each layer, a basic linguistic information unit is identified; basic units are separated but still interlinked each other across the different layers. The basic conceptual components of the MILE lexical model are the following:

- a) the MILE Lexical Classes (MLC) represent the main building blocks which formalize the basic lexical notions. They can be seen as a set of structural elements organized in a layered fashion: they constitute an ontology of lexical objects as an abstraction over different lexical models and architectures. These elements are the backbone of the structural model. These include main syntactic constructions, basic operations and conditions to establish multilingual links, macro-semantic objects, such as lexical conceptual templates acting as general constraints for the encoding of semantic units.
- b) the MILE Lexical Data Categories (MDC) which constitute the attributes and values to adorn the structural classes and allow concrete entries to be instantiated. Typical instances of MDCs are syntactic and semantic features, semantic relations, syntactic constructions, predicates and arguments etc.

MILE appears especially suited to our needs by virtue of being a) modular (different levels independently encoded), and b) granular (different degrees of depth at which an entry can be described at each level).

Since our case study concerns the semantic information of a lexical entry, we will concentrate on the

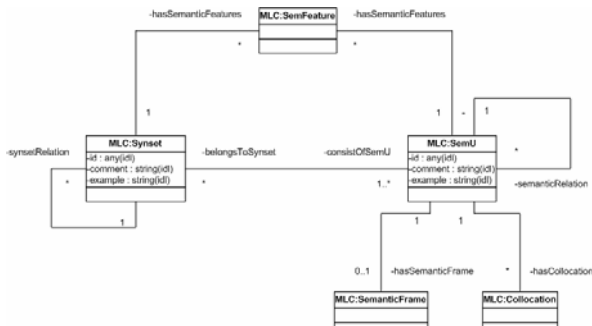


Figure 1: MILE semantic layer

semantic layer only, as illustrated in Figure 1.

Originally, in order to meet expectations placed upon lexicons as critical resources for content processing in the Semantic Web, the MILE syntactic and semantic lexical objects have been formalized in RDF(S), thus providing a web-based means to implement the MILE architecture and allowing for encoding individual lexical entries as instances of the model (Ide *et al.*, 2003; Bertagna *et al.*, 2004). In the framework of our project, by situating our work in the context of W3C standards and relying on standardized technologies underlying this community, the original RDF schema for ISLE lexical entries has been made compliant to OWL.

4. Integrating lexicons using LeXFlow

LeXFlow is not to be intended as a tool for the compilation or editing of lexicons (although it can be used to such an end). While different flows can be envisaged,

depending on the particular needs as well as on the particular attitude towards the work of a lexicographer, we demonstrate the potential of LeXFlow by illustrating its application to the case where two different semantic lexicons interact by reciprocally enriching themselves and integrating information coming from corpora. To this end, we have designed a sample lexical flow (see Figure 2) corresponding to the scenario where an entry of a lexicon A becomes enriched via basically two steps. First, by virtue of being mapped onto a corresponding entry belonging to lexicon B, the entry inherits the semantic relations available in lexicon B, and vice-versa. Second, by resorting to an automatic application that acquires information about semantic relations from corpora, the relations acquired are integrated into the entry and proposed to the human encoder for final checking and validation. The aim of this lexical flow is thus threefold:

- a) to enrich the entries of a lexicon with information coming from corpora and from an external lexicon;
- b) to show how the MILE lexical model not only allows, but enforces the integration;
- c) to provide an instrument, based on the MILE model, that allows the creation of enriched lexical entries, where the information coming from different lexicons is fused.

For our purposes, we chose to rely on the SIMPLE/CLIPS (Ruimy *et al.* 2003) and the ItalWordNet (Roventini *et al.* 2003) lexicons. These two semantic lexicons represent two very different attitudes towards the description of semantic content, and hence encode different types of information. In our scenario, it is assumed that the two lexicons are already represented according to the MILE specifications.

We recall that, according to the MILE model, an entry coincides with a given sense of a word (a SemU, Semantic Unit). In the simplified MILE-conformant entry schema we have adopted, each SemU is encoded as a single document. A SemU is described by means of the following attributes: a) an ID b) a gloss c) the lemma d) an optional example and e) an indication of the source. For the sake of readability, moreover, we overtly simplified the complexity of the two lexicon encodings by concentrating only on a subset of the range of semantic information available and actually encoded in lexicons. In particular, we decided to focus on the bunch of semantic relations (hyponymy, synonymy, meronymy, and the like) that a given sense of a lexical entry has with other senses of the same lexicon. Thus, for the SIMPLE/CLIPS lexicon, each SemU is further described by means of a list of semantic relations, each of them linked to a target SemU. On the other hand, in the MILE-conformant version of the ItalWordNet lexicon, each SemU corresponds to a variant of a given synset. Apart from the general descriptive fields described above, a wordnet-derived SemU only contains indication of the native synset, a notion expressed by the “belongsToSynset” relation. The semantic relations describing the relational context of a variant are described inside the synset.

In the following subsections we give a step-by-step description of the flow, whose overall picture is represented in Figure 2.

The figure clearly illustrates the different agents participating to the flow. Rectangles represent human actors over the entries, while the other figures symbolize

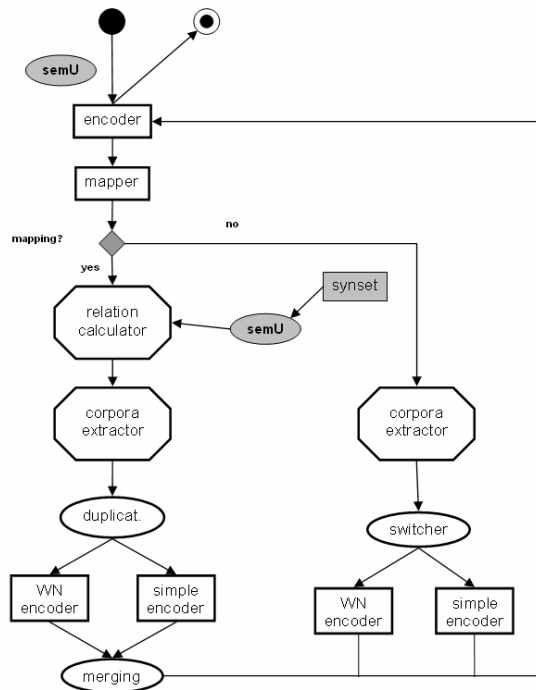


Figure 2: Lexical flow activity diagram

software agents: ovals are internal agents and octagons external ones.

4.1. Starting the flow: the mapping phase

In this scenario, a user or *encoder* starts by selecting an entry of a semantic lexicon that will represent the instance to be processed by the flow. Suppose that the selected entry is the SemU “car₁”, belonging to the SIMPLE/CLIPS lexicon. After this first step, the entry becomes processed by another user, having the role of “mapper”. The mapper selects a corresponding entry belonging to the ItalWordNet lexicon that expresses the same sense. Let’s assume that the mapper has identified a corresponding entry in the SemU “car₂” belonging to the Synset “car₂_auto₁_machine₄” of the ItalWordNet lexicon. For the sake of simplicity we hypothesize a human agent, but the same role could be performed by a software agent. To this end, we are investigating the possibility of automatically exploiting the procedure described in (Ruimy & Roventini, 2005).

4.2. Merging of semantic relations

If the mapping procedure is successful, then the two instances (entries) are loaded and aggregated in a single object. At this stage, this new object includes all the relations originally pertaining to the originating instances. That is, in this new object there will be the semantic relations as expressed in the SIMPLE-CLIPS lexicon as well as the Synset Relations as expressed in the IWN lexicon. The two different types of semantic relations will target the original targets, that is, the original SemUs for the SIMPLE lexicon and the original synsets for the IWN lexicon.

The following step is represented by the “relation calculator”. This software agent is responsible for creating for each lexicon a set of candidate relations on the basis of those available in the other lexicon. It does so by

performing two operations: first, it translates the semantic relations coming from a lexicon into the parlance of the other lexicon. Second, it creates for the imported relations as many candidate targets as are the original targets (either SemUs or Synsets). For instance, let’s suppose that the SIMPLE entry for “car₁” has a “has_as_part” semantic relation with another entry, namely “wheel₁”. Figure 3 illustrates this scenario.

The Relation Calculator then creates a translation of

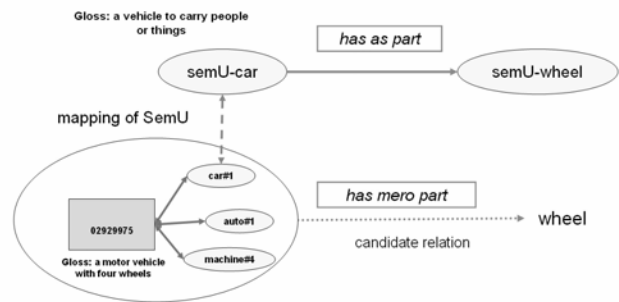


Figure 3: Candidate synset relations

each semantic relation into the language of the other lexicon, to be proposed for validation in a subsequent step. In the case at hand, it will translate the “has_as_part” relation into the corresponding “has_mero_part” synset relation. The targets of these candidate relations will not be SemU, but the procedure will propose a candidate lemma for each relation. It will be the encoder’s duty to associate a proper SemU belonging to his lexicon to a candidate relation. Moreover, if the SIMPLE-derived SemU contains some “has-synonym” relations then LeXFlow proposes a widening of the IWN synset by means of the lemma corresponding to the target SemU. On the other hand, for each synset relation encoded for a WordNet-derived SemU for which there is an equivalent

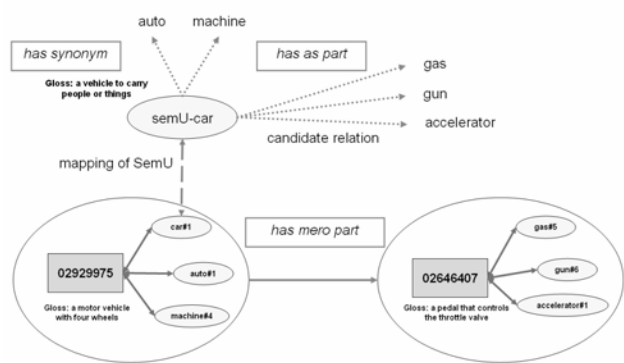


Figure 4: Candidate semantic relations

relation in the SIMPLE parlance, LeXFlow proposes as many candidate semantic relations as the SemUs contained in the target synset (see Figure 4).

Once again, every candidate semantic relation points to a lemma. In addition, LeXFlow creates as many semantic relations of the “has_synonym” type as are the variants belonging to the IWN corresponding synset.

The Relation Calculator simply ignores those relations that cannot be mapped.

4.3. Automatic acquisition from corpora

At this stage, the instance representing the unit under processing by the flow has been enriched with a set of potential semantic relations, as a result of the cross-breeding between the corresponding entries as encoded in the two source lexicons. The following step is represented by the action of an application that acquires information about *part-of* relations by identifying syntactic constructions in a vast Italian corpus of about 90 million words (Marinelli *et al.*, 2003). The corpus was previously analysed by Chunk-It (Lenci, Montemagni & Pirrelli, 2003), a chunker developed at ILC-CNR as part of a complete chain for the linguistic analysis of Italian. The flow invokes the application by sending a query on the basis of the lemma of the entry under processing. The application essentially consists in a grammar whose rules are syntactic patterns that can be indicative of meronymy relations. The output of the automatic procedure is then acquired by LeXFlow, that takes care of creating the appropriate candidate semantic and synset relations for each lemma that is proposed by the application. A lemma is automatically discarded as a candidate target if it is already present as target of a semantic or synset relation in the list of those already encoded in the entry (either originally or as a result of the merging step).

4.4. Enrichment of semantic relations

After these steps, LeXFlow duplicates the instance and sends it to two human agents, identified as a SIMPLE-encoder and an IWN encoder. Their duty consists in accepting or discarding the proposed relations, as well as choosing the appropriate target SemUs or Synset for each relation that is proposed. It is worth noting that LeXFlow produces two separate views of the same enriched entry by showing only the portions that are relevant to the different starting lexicons. In other words, the SIMPLE encoder will be able to validate only the semantic relations already translated into the SIMPLE parlance. On the other hand, these will remain opaque to the IWN encoder, who will check the proposed Synset relations only. In each separate view, LeXFlow provides to the encoders a window where starting from the proposed target lemmas the user can either choose the target SemU or Synset from the original lexicons (if already available), or either creating it from scratch.

4.5. Ending the flow

After the validation phase, the flow again makes a merging of the two versions of the entry, by joining the portions that have been modified by the two encoders. The merged entry is then returned to the initial user for a final check. If accepted, this new entry replaces the original entry in the lexical database. It is worth noting that the replacement takes place in both lexicons, thus providing a true contamination of the two worlds, although controlled. Since the entry is expressed in the MILE model, that provides the expressive power to allow for different views over the same semantic space, the contamination is not only allowed but enforced, thus paving the way for a truly merged lexicon to be created. In fact, the lexical flow described provides all the means for linking two lexicons in an integrated repository, with all entries opening doors over the two originating worlds. Initially the two lexical repositories are completely separated, although

compatible thanks to the interlingua provided by the MILE encoding. This situation is illustrated by Figure 5.

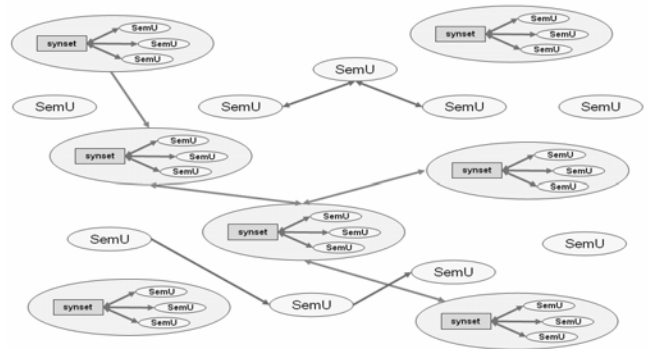


Figure 5: Lexicon initial state

SIMPLE SemUs and IWN SemUs co-exist into the same space, but are by no means connected, with the former

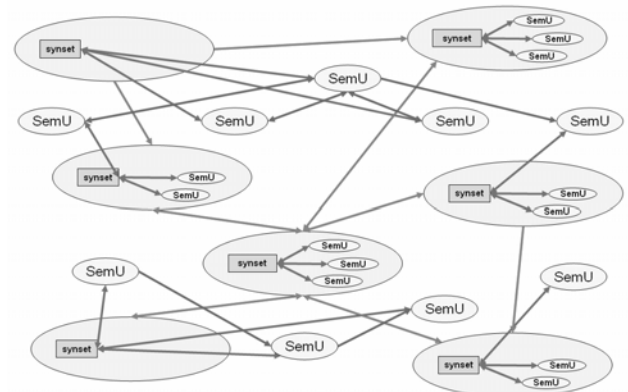


Figure 6: Lexicon running state

being linked only among themselves, and the latter only living into the restricted space of the Synsets to which they belong. After completing several flows, we gradually arrive at a situation where the two lexicons begin to integrate, with cross-breeded SemUs (participating of the properties of both lexicons) throwing links to IWN synsets and to SIMPLE SemUs (see Figure 6).

5. Concluding remarks and future works

Following its architectural principles, LeXFlow allows to distinguish among the flow engine and the description of the particular flow to be instantiated. Accordingly, many different flows can be envisaged from the one described in this paper. By simply describing different agents and roles, LeXFlow can be applied to the compilation of new lexical entries, to the quality control and checking, or to the creation of entries by combining information coming from different sources. In this paper we have illustrated an application of LeXFlow to the merging of different semantic lexicons, with a focus on the enrichment of source lexicons. The same principles can be applied in a scenario where a user is interested in combining different layers of lexical information, for instance phonetic and morphological information (see Monachini *et al.*, in press).

In the flow described in this paper the outcoming entries enter again into the original lexical repositories, and their merging is almost exclusively exploited in order to enrich their respective set of semantic relations. However, the new entries potentially contain the seeds for representing the building blocks of a truly integrated lexicon, where all the entries are in common. Investigating the possibility of creating a new global lexicon, where each addition or deletion of entries on each side (SIMPLE or IWN) has immediate and automatic consequences on the other represents the commitment of our future work.

6. References

- Bertagna, F., Lenci, A., Monachini, M., Calzolari, N. (2004). Content Interoperability of Lexical Resources: Open Issues and MILE Perspectives. In *Proceedings of the Fourth International Conference on Language Resources and Evaluation*. Paris, The European Language Resources Association (ELRA), pp. 131-134.
- Calzolari, N., Bertagna, F., Lenci, A., Monachini, M. (eds.) (2003). *Standards and Best Practice for Multilingual Computational Lexicons. MILE (the Multilingual ISLE Lexical Entry)*. ISLE CLWG Deliverable D2.2 & 3.2. Pisa.
- Calzolari, N., Lenci, A., Zampolli, A. (2001). International Standards for Multilingual Resource Sharing: the ISLE Computational Lexicon Working Group. In *Proceedings of the ACL-EACL Workshop on Sharing Tools and Resources*, Toulouse, CNRS, pp. 71-78.
- Calzolari N., Soria C. (2005). A new paradigm for an Open Distributed Language Resource Infrastructure: the case of Computational Lexicons. In *Proceedings of the AAAI Spring Symposium "Knowledge Collection from Volunteer Contributors (KVC05)"*, Stanford, CA, March 21-23.
- Francopoulo, G., George, M., Calzolari, N., Monachini, M., Bel, N., Pet, M., Soria, C. (in press). Lexical Markup Framework (LMF). Accepted for publication in *Proceedings of LREC2006*, Genoa, Italy.
- Gulrajani G. (2002). SHAWEL: Sharable and Interactive Web-Lexicons. Paper available at: <http://emeld.org/workshop/2002/presentations/wittenburg/shawel-paper-final.doc>.
- Ide, N., Lenci, A., Calzolari, N. (2003). RDF Instantiation of ISLE/MILE Lexical Entries. In *Proceedings of the ACL 2003 Workshop on Linguistic Annotation: Getting the Model Right*, Sapporo (Japan), ACL, pp. 25-34.
- ISO 12620 - Terminology and other language resources – Data Categories – Data category selection for electronic lexical resources. ISO/TC37/SC4.
- Lenci A., Montemagni S., Pirrelli V. (2003). *Chunk-it*. An Italian shallow parser for robust syntactic annotation. In A. Zampolli, N. Calzolari, L. Cignoni, (eds.), *Computational Linguistics in Pisa – Linguistica Computazionale a Pisa. Linguistica Computazionale*, Special Issue, XVI-XVII. Pisa-Roma, IEPI, vol. I, pp. 353-386.
- Marchetti, A., Minutoli, S., Lazzareschi, P., Martinelli, M. (2001). A System for Managing Documents in a Step by Step Process. In *Proc. XML World Euro Edition*, 26-28 March 2001, Amsterdam-Holland.
- Marchetti, A., Tesconi, M., Minutoli, S. (2005). XFlow: An XML-Based Document-Centric Workflow. In *Proceedings of the Sixth International Conference on Web Information Systems Engineering (WISE)*, New York, NY, USA, pp. 290 - 303.
- Marinelli R., Biagini L., Bindi R., Goggi S., Monachini M., Orsolini P., Picchi E., Rossi S., Calzolari N., Zampolli A. (2003). The Italian PAROLE corpus: an overview. In Zampolli A., Calzolari N., Cignoni L. (eds.), *Computational Linguistics in Pisa*, Special Issue of *Linguistica Computazionale*, Vol. XVIII-XIX, Istituto Editoriale e Poligrafico Internazionale, Pisa-Roma, 2003.
- Monachini, M., Calzolari, N., Choukri, K., Friedrich, J., Mammini, M., Odijk, J., Ulivieri, M. (in press). Unified lexicon and unified morpho-syntactic specifications for written and spoken Italian. Accepted for publication in *Proceedings of LREC2006*, Genoa, Italy..
- Romary L., Francopoulo G., Monachini M., Salmon-Alt S. (in press). Lexical Markup Framework (LMF): working to reach a consensual ISO standard on lexicons. Accepted for publication in *Proceedings of LREC2006*, Genoa, Italy.
- Roventini A., Alonge A., Bertagna F., Calzolari N., Girardi C., Magnini B., Marinelli R., Speranza M., Zampolli A. (2003). ItalWordNet: Building a Large Semantic Database for the Automatic Treatment of Italian. In Zampolli A., Calzolari N., Cignoni L. (eds.), *Computational Linguistics in Pisa, Special Issue of Linguistica Computazionale*, Vol. XVIII-XIX, Istituto Editoriale e Poligrafico Internazionale, Pisa-Roma.
- Ruimy N., Monachini M., Gola E., Calzolari N., Del Fiorentino M.C., Ulivieri M., Rossi S. (2003). A Computational Semantic Lexicon of Italian: SIMPLE. In Zampolli A., Calzolari N., Cignoni L. (eds.), *Computational Linguistics in Pisa, Special Issue of Linguistica Computazionale*, Vol. XVIII-XIX, Istituto Editoriale e Poligrafico Internazionale, Pisa-Roma, 2003.
- Ruimy N., Roventini A. (2005). Towards the linking of two electronic lexical databases of Italian. In *Proceedings of L&T'05 - Language Technologies as a Challenge for Computer Science and Linguistics*, Poznan, Poland, pp. 230-234.
- Sanfilippo, A., Calzolari, N., Ananiadou, S., Gaizauskas, R., Saint-Dizier, P., Vossen, P. (eds.) (1999). EAGLES Recommendations on Semantic Encoding. *EAGLES LE3-4244 Final Report*. (Also URL: <http://www.ilc.cnr.it/Eagles96/rep2>).
- Wittenburg, P., Kemps-Snijders, M. (2006). *Some LIRICS Topics*, 2006 (available at http://lirics.loria.fr/doc_pub/lirics-barca.ppt#480,2,LMF_Topics)