

SIMPLE: A General Framework for the Development of Multilingual Lexicons

Nuria Bel³, Federica Busa^{1,7}, Nicoletta Calzolari¹, Elisabetta Gola¹, Alessandro Lenci^{1,2}, Monica Monachini¹, Antoine Ogonowski⁶, Ivonne Peters⁵, Wim Peters⁵, Nilda Ruimy¹, Marta Villegas^{3,4}, Antonio Zampolli^{1,2}

¹ Istituto di Linguistica Computazionale, Pisa

² Università di Pisa

³ Gilc - Universitat de Barcelona

⁴ Institut d'Estudis Catalans

⁵ University of Sheffield

⁶ LexiQuest

⁷ Brandeis University

Abstract

The project LE-SIMPLE is an innovative attempt of building harmonized syntactic-semantic lexicons for 12 European languages, aimed at use in different Human Language Technology applications. SIMPLE provides a general design model for the encoding of a large amount of semantic information, spanning from ontological typing, to argument structure and terminology. SIMPLE thus provides a general framework for resource development, where state-of-the-art results in lexical semantics are coupled with the needs of Language Engineering applications accessing semantic information.

1. Introduction

SIMPLE is a large project sponsored by EC DGXIII in the framework of the Language Engineering programme. This project represents an innovative attempt to develop wide-coverage semantic lexicons for a large number of languages (12),¹ with a harmonised common model that encodes structured "semantic types" and semantic (subcategorization) frames. Even though SIMPLE is a lexicon building project, it also addresses challenging research issues and provides a framework for testing and evaluating the maturity of the current state-of-the-art in the realm of lexical semantics grounded on, and connected to, a syntactic foundation.

Many theoretical approaches are currently tackling different aspects of lexical semantics. However, such approaches have to be tested *i*) with wide-coverage and multilingual implementations, and *ii*) with respect to their actual usability and usefulness in real-world systems both of mono- and multi-lingual nature. The SIMPLE project addresses point *i*) directly, while providing the necessary platform to allow application projects to address point *ii*), thus contributing to the need of a core set of language resources for the EU languages.

SIMPLE should be considered as a follow up to the LE-PAROLE project (Ruimy *et al.*, 1998) because it adds a semantic layer to a subset of the existing morphological and syntactic layers developed by PAROLE. The semantic lexicons (about 10,000 word meanings) are built in a uniform way for the 12 PAROLE languages. These lexicons are partially corpus-based, exploiting the harmonised and representative corpora built within PAROLE. The lexicons are designed bearing in mind a future cross-language linking: they share and are built around the same *core ontology* and the same set of *semantic templates*. The "base concepts" identified by EuroWordNet (about 800 senses at a high level in the

taxonomy) were used as a core set of senses, so that a cross-language link for all the 12 languages is already provided automatically through their link to the EuroWordNet Interlingual Index (Vossen *et al.*, 1998).

2. The model

In the first stage of the project, the formal representation of the "conceptual core" of the lexicons was specified, i.e. the basic structured set of "semantic types" (the *SIMPLE ontology*) and the basic set of notions to be encoded for each sense. The development of 12 harmonised semantic lexicons requires strong mechanisms for guaranteeing uniformity and consistency. The multilingual aspect translates into the need to identify elements of the semantic vocabulary for structuring word meanings that are both language independent and able to capture linguistically useful generalisations for different NLP tasks.

The SIMPLE model is based on the recommendations of the EAGLES Lexicon/Semantics Working Group (Sanfilippo *et al.*, 1998) and on extensions of Generative Lexicon theory (cf. Pustejovsky, 1998; Busa *et al.*, 1999). An important part of the background of SIMPLE is also represented by the two ACQUILEX projects (Calzolari, 1991) and the DELIS project (Monachini *et al.*, 1994), especially in connection with the techniques developed for sense extraction and integration into lexical knowledge bases. An essential characteristic of the Generative Lexicon is its ability to capture the various dimensions of word meaning. The basic vocabulary relies on an extension of "Qualia Structure" (cf. Pustejovsky, 1995) for structuring the semantic/conceptual types as a representational device for expressing the multi-dimensional aspect of word meaning. The model has a high degree of generality in that it provides the same mechanisms for generating broad-coverage and coherent concepts for different semantic areas (e.g. entities, events, abstract nouns, etc.).

Besides some local aspects of novelty (e.g. a refinement of the Qualia organisation of semantic

¹ Catalan, Danish, Dutch, English, Finnish, French, German, Greek, Italian, Portuguese, Spanish, Swedish.

information, taking into account also applicative requirements), the real innovation and the strength of the project design lies *i*) in the thoroughness of description, covering many different semantic aspects (often dealt with separately in existing lexicons), and in the choices done in their combination in a global model; *ii*) in the application of the same rich model to so many languages of different type (spanning from Romance languages, to Germanic ones and to Finnish); *iii*) in establishing a common methodology of building all the lexicons in a peculiar combination of top-down and bottom-up strategies; *iv*) in the possibility of verifying a number of theoretical claims on a large number of entries and for a variety of different languages.

In order to combine the theoretical framework with the practical lexicographic task of lexicon encoding, SIMPLE has created a common "library" of language independent *templates*, which act as "blueprints" for any given type - reflecting the conditions of well-formedness and providing constraints for lexical items belonging to that type. The relevance of this approach for building consistent resources is that types both provide the formal specifications and guide subsequent encoding, thus satisfying theoretical and practical methodological requirements.

The SIMPLE model contains three types of formal entities:

- **Semantic Units** - word senses are encoded as *Semantic Units* or *SemU*. Each SemU is assigned a *semantic type* from the Ontology, plus other sorts of information specified in the associated *template*, which contribute to the characterisation of the word sense.
- **Semantic Type** - it corresponds to the semantic type which is assigned to SemUs. Each type involves structured information, organised in the four Qualia Roles, adopted in the Generative Lexicon framework. The Qualia information is sorted out into *type-defining information* and *additional information*. The former is information that intrinsically defines a semantic type as it is. In other words, a SemU can not be assigned a certain type, unless its semantic content includes the information that defines that type. On the other hand, additional information specifies further components of a SemU, rather than entering into the characterisation of its semantic type.
- **Template** - a schematic structure which the lexicographer uses to encode a given lexical item. The template expresses the semantic type, plus other sorts of information. Templates are intended to guide, harmonise, and facilitate the lexicographic work. A set of top templates have been prepared during the specification phase, while more specific ones will be eventually elaborated by the different partners according to the need of encoding more specific concepts in a given language.

The SIMPLE model provides the formal specification for the representation and encoding of the following information: semantic type, corresponding to the template the SemU instantiates; domain information; lexicographic gloss; argument structure for predicative

SemUs; selectional restrictions on the arguments; event type, to characterise the aspectual properties of verbal predicates; link of the arguments to the syntactic subcategorization frames, as represented in the PAROLE lexicons; Qualia structure; information about regular polysemous alternation in which a word sense may enter; information concerning cross-part of speech relations (e.g. *intelligent - intelligence; writer - to write*).

The semantic types in SIMPLE form a general *Ontology*, which is structured in such a way to take into accounts the principles of *orthogonal organisation* of types, as formalised in the Generative Lexicon (Pustejovsky, 1995).

The hierarchy of types has been further subdivided into two layers:

- **The Core Ontology** - it is formed by those types which have been identified as the central and common ones for the construction of the different lexicons in SIMPLE, and which represent the highest nodes in the hierarchy of types.
- **Recommended Ontology** - this is formed by more specific types (lower nodes in the hierarchy), which provide a more granular organisation of the word-senses.

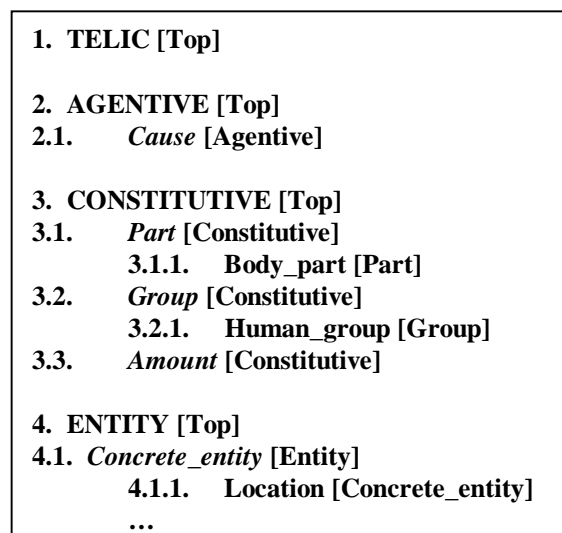


Fig.1 - The SIMPLE ontology. A sample

As illustrated in fig. 1, principles of Qualia Structure in the Generative Lexicons have also been adopted to organize the top-level ontology. The type **Constitutive**, for instance, dominates those semantic types describing word senses (such as part, constituent, element) whose semantic contribution is fully determined only by meronymic relations with other SemUs (since hyperonymic links are in these cases quite uninformative). This solution has proven to be quite useful to provide a rich representation for SemUs belonging to areas of the lexicon (e.g. relational nouns, abstracts, etc.) that are notoriously quite resistant to be captured in semantic type systems.

3. The multifaceted expressive resources in SIMPLE

It is widely acknowledged that formal ontologies represent a useful resource to characterise the semantic

content of a word, thus providing input for different sorts of NLP applications that need to access the conceptual content of words. Actually, a growing number of systems of Information Extraction, Information Retrieval, Word Sense Disambiguation, etc. include conceptual taxonomies, in terms of which word senses are classified. Both on the theoretical and on the applicative side, therefore, a wide range of often competing ontologies have been created: many of them are also tailored to meet particular "knowledge management" requirements of specific domains and purposes. On the other hand, building an ontology of types for a general purpose lexicon surely represents a problem with a bigger order of complexity: different areas of the lexicon to be covered, establishing *domain independent* criteria for the selection of the relevant types, etc. A general purpose resource like SIMPLE must also face the problem that various potential users of the resource might need to carve out different parts of the lexicon, and to extend them to meet their needs. Extensions could concern both the *size* of the resource and the *granularity* of the semantic information which is encoded; that is to say users might be interested in adding more specific senses, as well as to add semantic information to the existing ones (e.g. for domain specific requirements). This means that SIMPLE has to provide a general framework for semantic encoding, which is able to:

1. facilitate the customisation of the resource;
2. allow an easy and fully consistent extension of different areas of the lexicon.

SIMPLE tries to comply with these requirements by providing a rich expressive language for the representation of semantic information, and by associating each type of the ontology with a well-specified cluster of information which defines the type itself. Thus, the template associated to a type provides a sort of *interpretation* of the type itself. The full expressive power of the SIMPLE model is given by a wide set of features and relations, which are organised along the four Qualia dimensions, Formal, Agentive, Constitutive and Telic. Features are introduced to characterise those attributes for which a closed and restricted range of values can be specified. On the other hand, relations between SemUs have been defined for those aspects of lexical meaning that cannot be easily reduced to a closed range of attribute-values pairs. Here is a small sample of the semantic relations in SIMPLE (cf. Lenci *et al.*, 1999):

Name	Description	Example	Type
<i>Is_a_member_of</i>	<SemU1> is a member or element of <SemU2>.	<senator>; <senate>	<i>Constitutive</i>
<i>Is_a_part_of</i>	<SemU1> is a part of <SemU2>	<head>;<b ody>	<i>Constitutive</i>
<i>Used_for</i>	<SemU1> is typically used for <SemU2>	<eye>;<see >	<i>Telic</i>
<i>Purpose</i>	<SemU2> is an event corresponding to the intended purpose of	<send>;<re ceive>	<i>Telic</i>

	<SemU1>		
--	---------	--	--

Relations are also organised along a taxonomic hierarchy, allowing for the possibility of underspecification, as well as the introduction of more refined subtypes of a given relation.

Templates provide the information that is type-defining for a given semantic type. Lexicographers can also further specify the semantic information in a SemU, by either adding other relations or features in the Qualia Structure, or by adding other types of information (e.g. domain information, collocations, etc.). Take, for instance, the template associated to the type **Instrument**:

UseM:	<u>1</u>
Template_Type:	[Instrument]
nification_path:	[Concrete_entity Artifact _{Agentive} Telic]
Domain:	<i>General</i>
Semantic Class:	<Nil>
Gloss:	//free//
Pred_Rep.:	<Nil>
Selectional Restr.:	<Nil>
Derivation:	<Nil>
Formal:	<i>isa</i> (1,<instrument>)
Agentive:	<i>created_by</i> (1,<UseM>: [Creation])
Constitutive:	<i>made_of</i> (1,<UseM>) //optional// <i>has_as_part</i> (1,<UseM>) //optional//
Telic:	<i>used_for</i> (1,<UseM>: [Event])
Synonymy:	<Nil>
Collocates:	<i>Collocates</i> (<UseM1>,...,<UseMn>)
Complex:	<Nil> //for regular polysemy//

This template describes the type **Instrument** as being inherently defined by agentive information (i.e. concerning the origin of an instrument), and telic information (i.e. what an instrument is used for), besides the standard hyperonymic relation.

In order to appreciate the peculiarities of the semantic representation in the SIMPLE model, it is interesting to compare it with other well-known semantic organizations, like for instance the one in WordNet 1.6 (Fellbaum, 1998). In WordNet, semantic lexical information is provided by a full, "verticalized" taxonomical hierarchy connecting a given synset to a top node. Thus, the backbone of the hierarchy (at least for nouns) is represented by the *isa* relation. For instance, the following is the WordNet 1.6 description of one of the senses of *lancet*:

Sense 2

```
lancet, lance
=> surgical knife
=> knife
=> edge tool
=> cutter, cutlery, cutting tool
=> cutting implement
=> tool
=> implement
```

```

=> instrumentality, instrumentation
=> artifact, artefact
=> object, physical object
=> entity, something

=> surgical instrument
=> medical instrument
=> instrument
=> device
=> instrumentality, instrumentation
=> artifact, artefact
=> object, physical object
=> entity, something

```

One well-known characteristic of this style of representation is that actually the nodes of the *isa* hierarchy refer to various and heterogeneous kinds of information. For instance, at the third step in the sense 2 for *lancet* ("a surgical knife with a pointed double-edged blade; used for punctures and small incisions"), we find information referring to a constitutive aspect of lancets ("edge tool"); two steps further, we instead find information referring to the purpose typically associated with lancets ("cutting implement"). Keeping on climbing up, we find information on the origin of lancets ("artifact"). Finally, other relevant pieces of information, such as for instance the fact that lancets belong to the domain of surgery, are also spread out in the taxonomy. Therefore, although the WordNet entry contains a rich amount of information characterizing the relevant sense of *lancet*, this information is not fully explicit, and is therefore not directly and easily accessible by applications. Moreover, different types of information do not have a "fixed" location within the *isa*-hierarchy, so that the same type of information (e.g. information concerning the typical purpose of an artifact or the material it is made of) might be located at different levels of the hierarchy for different entries. This fact surely represents another source of potential difficulty for those applications that need or want to target specific pieces of semantic information.

Differently from this approach to semantic representation, SIMPLE, following the main tenets of the Generative Lexicon, sorts out the various types of information entering into the characterization of a given word sense, as it can be seen in the above template for **Instruments**. Moreover, each piece of semantic information is also typed and inserted into structured hierarchies, each explicitly characterizing a certain aspect of the semantic content of nouns, verbs and adjectives. This way, the semantic information identifying word senses is fully explicit, and can directly and selectively be targeted by NLP applications. Finally, differently from WordNet-style architectures, lexical information in SIMPLE is structured in terms of small, local semantic networks, which operate in combination with feature-based information and a rich description of the argument structure and selectional preferences of predicative entries.

The following is the SemU for the above mentioned sense of *lancet*, instantiating the template **Instrument**:

UseM:	Lancet
BC number:	
Template Type:	[Instrument]
Unification path:	[Concrete_entity] Artifact _{Agentive} Telic]
Domain:	Medicine
Semantic Class:	Instrument
Gloss:	a surgical knife with a pointed

	double-edged blade; used for punctures and small incisions
Pred_Rep.:	<Nil>
Selectional Restr.:	<Nil>
Derivation:	<Nil>
Formal:	<i>isa</i> (<lancet>, <knife>: [Instrument])
Agentive:	<i>created_by</i> (<lancet>, <make>: [Creation])
Constitutive:	<i>made_of</i> (<lancet>, <metal>: [Substance]) <i>has_as_part</i> (<lancet>, <edge>: [Part])
Telic:	<i>used_for</i> (<lancet>, <cut>: [Constitutive_change]) <i>used_by</i> (<lancet>, <doctor>)
Synonymy:	<Nil>
Collocates:	<i>Collocates</i> (<UseM1>, ..., <UseMn>)
Complex:	<Nil>

It is important to notice that the Qualia information of the SemU is formed by the relations "inherited" by the template the SemU instantiates, plus other additional information. The former type is - so to speak - what defines a lancet as being of the type **Instrument**.

Another advantage of this solution is that it is possible to capture the different semantic loads of different classes of word senses, by calibrating the usage of the different pieces and types of information made available by the model.

The wide range of types of information by means of which lexical content is captured in SIMPLE also makes the lexicon a more versatile tool for Language Engineering, trying to meet some of growing needs of NLP applications. Actually, it is widely proven that crucial NLP tasks (IE, WSD, NP Recognition, etc.) need to access multidimensional aspects of word meaning. For instance, the proper identification of the semantic contribution of a NP requires to access a very rich representation of the semantic content of the nominal heads. Actually, it is the sense of the nominal head that determines the semantic relation expressed by a modifying PP. Take for instance the following expressions:

- (1)
- a. *la pagina del libro*
'the page of the book'
 - b. *il difensore della Juventus*
'the Juventus fullback'
 - c. *il suonatore di liuto*
'the liute player'
 - d. *il tavolo di legno*
'the wooden table'

In (1a), the noun head and the PP are in a *part_of* relation that can be easily identified given a sufficiently rich representation of the relevant sense of *pagina* (page), containing for instance a proper meronymic relation with books and other semiotic artifacts. On the other hand the same syntactic pattern is rather to be interpreted in (1b) as expressing a *member_of* relation between the noun and the PP modifier. Again, the lexicon can have a crucial role in identifying it, for instance specifying in the lexical entry for the relevant sense of *difensore* (fullback)

fullbacks are member of football teams. As for (1c) and (1d), the correct identification of the semantic content of the whole NP requires the identification respectively of the "telic" relation between the musical instrument and its player, and of the fact that the PP *di legno* expresses the matter out of which table might be composed.

Besides, notice that Qualia-like information defining the semantic content of a certain word sense must also be combined with information concerning the predicative structure of word senses. Take for instance the following case:

- (2) a. *il difensore di Clinton*
'Clinton's defender'
b. *il difensore della Juventus*
'the Juventus fullback'

The word *difensore* actually has two senses, one corresponding to the English *defender* (SemU₁), and the latter to the English *fullback* (SemU₂). The interesting fact is that only the former sense is predicative (actually deriving its argument structure from the verb *difendere*, to which *difensore* is morphologically related). The particular argument structure and selectional preferences of SemU₁, combined with Qualia information, have a crucial role in guiding the disambiguation of the word *difensore*, thereby providing the correct interpretation of NPs like those in (2). Thus rich lexical resources, which are able to tackle simultaneously different, but equivalently crucial aspects of word meaning, appear to have a crucial role to enhance the performance of NLP systems.

The SIMPLE framework has many advantages:

1. Thanks to the different types of semantic information that can be represented in SIMPLE, the model is geared to customisation for specific needs. Possible extensions of the lexicon may thus target peculiar aspects of the semantic content (e.g. by using more specific relations), without losing the general consistency of the system.
2. It allows a high degree of underspecification in type assignment, which is extremely useful in the phase of lexicon construction (especially in multilingual environment), in order to maximise the consistency of the encoding. Actually, the problems of applying whatever system of semantic types in semantic encoding are well-known: assuming a system of semantic types means to commit oneself to a particular conceptualisation of reality that is in many cases unable to fully capture lexical richness. Besides, in many cases it is difficult to provide firm criteria for the selection of a given semantic type. The usual solution for lexicographers is *underspecification*, i.e. recurring to the highest nodes in a taxonomy. This has the obvious shortcoming of generating quite uninformative representations. SIMPLE addresses this problem by the combined action of template-assignment and the possibility of adding other optional information taken by the list of available relations and features. In other terms, it is possible to assign an underspecified type to a SemU, without losing the possibility of expressing important parts of its semantic contribution. Therefore, SIMPLE allows

recurring to *type-underspecification*, without losing in *informativeness*.

New types and templates can be created, by selecting particular pieces of information out of sets of semantically homogenous SemUs. It is thus possible to customise the lexicon and the type system both for application/domain-specific needs and to capture language-specific peculiarities.

3.1. The representation of polysemy

A well-known problem of lexical semantics is the representation of polysemy. Although SIMPLE cannot but scratch the surface of the problem, it provides a way to represent the distinction between purely ambiguous readings of a word and (regular) polysemous senses. Polysemous senses are connected through a *polysemous relation* between the corresponding SemUs. A starting set of regular polysemous classes has been selected (corresponding to the best-known polysemous alternations, e.g. Building-Institution, Animal-Food, etc.), and for each of them a specific relation has been introduced. For instance, the name *school* has at least two SemUs, <school₁>, meaning "building which is used for educative purposes", and <school₂>, meaning "educational institution". The fact that *school* is a polysemous word, and its senses belong to the regular *building-institution* alternation is represented by linking the two SemUs through the following relation:

BuildingInstitution (<school₁>, <school₂>)

On the other hand, purely ambiguous senses of a word (e.g. *bank*) are represented as SemUs that are not linked by any polysemous relation.

Individuating new polysemous classes is part of the research purposes of SIMPLE. One of the aims of the project is actually to carry out a campaign of *polysemy mining* among the various partners of the project. The target is to have a more detailed map of regular polysemy in various European languages, as well as to have an estimation of the language specific variation in this phenomenon.

3.2. Arguments and predicates in SIMPLE

One of the major efforts of SIMPLE is to encode the argument structure of verbs, adjectives and predicative nouns. Besides the number of semantic arguments and their selectional restrictions, for each predicative SemU the link of the arguments to the corresponding syntactic construction in the PAROLE lexicon is provided. Thus, SIMPLE also represents a powerful framework for the description of semantic argument frames, of selectional constraints, as well as of the syntactic realisations of arguments.

The issue of the syntax-semantics lexical interface is tackled in SIMPLE with three modules of information: *i*) the *syntactic layer* (in the PAROLE lexicon), containing the syntactic subcategorization frame of a word; *ii*) the *predicative layer* of the SemU, describing its argument structure; *iii*) the *correspondence layer*, where arguments are linked to syntactic positions, and a large array of constraints can be set, to express particularly complex links. The representation of *nominalizations* in SIMPLE is an interesting case of interplay among these modules.

For instance, the SemUs of the verb *destroy* and of the derived noun *destruction* are assigned the same predicative layer, i.e. they share the same abstract predicate (e.g. `PREDDestroy`) and the same argument structure. On the other hand, the patterns of surface realisation of arguments in the verb and the noun are captured at the correspondence layer, by establishing specific links with positions in the descriptions associated with the SemUs in the syntactic layer. Therefore, SIMPLE allows to easily capture the semantic similarity between verbs and derived nouns (e.g. that the selectional restrictions of the arguments are the same), and at the same time to describe their differences in the linking with syntax (e.g. with respect to the optional realisation of the arguments).

4. Conclusions

SIMPLE provides a general framework for the development, customisation and extension of "state-of-the-art" lexical resources. The project also consists in the construction of harmonised lexicons for 12 European languages. The range of information represented in SIMPLE makes these lexicons a necessary complementation to other existing resources, such as for instance those developed in the EuroWordNet project. All the lexical information is encoded in SGML and the whole SIMPLE model is fully represented according to the GENELEX DTD specifications (GENELEX, 1994); the lexicons will be made publicly available by ELRA together with the morphological and syntactic units. This way, the PAROLE-SIMPLE package will represent an integrated *lexical suite*, ready to be directly accessed by different sorts of NLP applications. External experts will validate the lexicons being built through sampling; validation of the ontology and type-system will also be carried out through its use for qualifying selectional restrictions.

The availability of high-quality lexicons will be crucial for the development of applications with semantic processing capability. SIMPLE relies on the existing EAGLES standards and, also through the enlargement in various National projects (e.g. Denmark, Italy, etc.), proposes itself as a *de facto* standard for computational lexical semantics, which will be developed towards a fully multilingual environment in the upcoming EU and NSF sponsored ISLE/EAGLES project.

5. References

Busa, F., Calzolari, N., Lenci, A. and J. Pustejovsky, 1999. Building a Semantic Lexicon: Structuring and Generating Concepts, paper presented at *The Third International Workshop on Computational Semantics*, 13-15 January 1999, Tilburg, The Netherlands.

Calzolari, N., 1991. *Acquiring and Representing Information in a Lexical Knowledge Base*, ILC-CNR, Pisa, ESPRIT BRA-3030/ACQUILEX - WP No. 16, March 1991.

Fellbaum, C. (ed.), 1998. *WordNet. An Electronic Lexical Database*, Cambridge, MA: The MIT Press.

GENELEX Consortium, 1994. *Report on the Semantic Layer*, Project EUREKA GENELEX, Version 2.1, September 1994.

Lenci, A. et al., 1999. *SIMPLE Work Package 2 - Linguistic Specifications*, Deliverable D2.1, October 1999, ILC-CNR, Pisa.

Monachini, M., Roventini, A., Alonge, A., Calzolari, N. and O. Corazzari, 1994. *Linguistic Analysis of Italian Perception and Speech Act Verbs*, ILC-CNR, Pisa, DELIS, Final Report, February 1994.

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

Pustejovsky, J., 1998. Specification of a Top Concept Lattice, ms. Brandeis University.

Ruimy, N., Corazzari, O., Gola, E., Spanu, A., Calzolari, N. and A. Zampolli, 1998. The European LE-PAROLE Project: The Italian Syntactic Lexicon, in *Proceedings of the First International Conference on Language resources and Evaluation*, Granada: 2141-248.

Sanfilippo, A. et al., 1998. *EAGLES Preliminary Recommendations on Semantic Encoding*, The EAGLES Lexicon Interest Group

Vossen, P., Bloksma, L., Rodriguez, H., Climent, S., Roventini, A., Bertagna, F., Alonge, A. and W. Peters, 1998. *The EuroWordNet Base Concepts and Top Ontology*, Deliverable D017, D034, D036, WP5, LE2-4003, 1998.