Procesamiento del Lenguaje Natural, Revista nº 44, marzo de 2010, pp 99-106 recibido 20-01-10 revisado 22-02-10 aceptado 03-03-10

Simple_PLUS: a network of lexical semantic relations

Simple_PLUS: una red de relaciones léxico-semánticas

Nilda Ruimy
Istituto di Linguistica Computazionale - CNR
Via Moruzzi, 1 – 56124 Pisa (Italy)
nilda.ruimy@ilc.cnr.it

Resumen: Este artículo trata de la base de datos léxico-semántica del italiano, *Simple_PLUS*, y particularmente de su núcleo central : la red de relaciones léxico-semánticas. Este recurso lexical tiene como base Parole-Simple-Clips, un léxico electrónico con cuatro niveles de descripción, elaborado según el modelo SIMPLE. *Simple_PLUS* se compone de 30.000 entradas semánticas, sean importadas del léxico fuente, sean recién creadas, todas dotadas de un amplio conjunto de información proporcionado por el modelo subyacente. En *Simple_PLUS*, aquella representación semántica fue enriquecida con una información relacional esencial, en un proceso semiautomático. Mas de 5.000 lazos que relacionan los eventos con sus participantes y los co-participantes entre sí ─ vínculos que no podían ser descritos antes por falta de medios de representación adecuados ─ fueron codificados mediante un vocabulario descriptivo apropiado, que fue prestado del modelo EuroWordNet. Estos lazos conceptuales, que enriquecen la representación predicativa del léxico, aportan un conocimiento lexical imprescindible para las tareas de PLN y la Web semántica.
Palabras clave: Léxico, relaciónes léxico-semánticas, modelo SIMPLE, modelo EuroWordNet

Abstract: The present article deals with the Italian lexical-semantic database *Simple_PLUS* and focuses on its essential core, i.e. the network of lexical semantic relations. This lexical resource builds on Parole-Simple-Clips, a four-layered electronic lexicon of Italian, founded on the SIMPLE model. *Simple_PLUS* consists of 30,000 semantic entries, partly imported from the source lexicon and partly newly created, but all encoding a wide-ranging set of information provided by the underpinning model. In *Simple_PLUS*, this semantic representation has been enriched with significant relational information, in a largely automated, inexpensive process. More than 5,000 relationships between events and their participants and among co-participants in events, links which were not capturable previously through lack of suitable representational means, have been encoded with the appropriate descriptive vocabulary borrowed from the EuroWordNet lexical model. Such conceptual links, which efficiently enhance the predicative representation in the lexicon, provide crucial lexical knowledge for NLP systems and for the Semantic Web.
Keywords: Lexicon, lexical semantic relations, SIMPLE model, EuroWordNet model

1 Introduction

Simple_PLUS is an enhanced subset of the PAROLE-SIMPLE-CLIPS (henceforth, PSC) four-layered computational lexicon of Italian language. The latter was developed over two EU-sponsored initiatives (MLAP/LE2-4017 PAROLE and LE4-8346 SIMPLE), and further extended and refined in the framework of the follow-up Italian project CLIPS (Corpora e Lessici dell'Italiano Parlato e Scritto). During this national project, the lexicon was enriched with a phonological representation of lexical units and the syntactic and semantic layers' coverage was extended, in part also by our partner, the Thamus Consortium in Salerno.

The part of the PSC lexicon elaborated by CNR-ILC in Pisa consists of 387,250 phonological entries, 53,000 morphological entries, 37,500 syntactic entries and 28,800 semantic entries, which provide a full-fledged description of lexical units at all linguistic levels. They implement the whole set of lexical

ISSN 1135-5948

information types offered by the underlying PAROLE-SIMPLE model (Ruimy et al., 1998; Lenci et al., 2000)[1], as well as some further refinements introduced in the context of the CLIPS project (Ruimy et al., 2003).

Simple_PLUS built on this ILC lexicon, undertaking an overall check and revision of the entries and enriching its level of semantic description with further relational information, with a view to gaining a sharper understanding of the semantic relationships holding among word senses in a sentence.

A minimal part of the additional information characterizing *Simple_PLUS* belongs to the very SIMPLE model: missing instantiations of SIMPLE relations were in fact introduced, and in particular synonymic links were strengthened. Most of the lexicon enhancement, however, consists in the addition of new conceptual links which were not capturable previously through lack of appropriate representational vocabulary in the SIMPLE model. The descriptive means for the representation of such links were borrowed from the EuroWordNet (EWN) model (Alonge, 1996; Vossen, 2002) which is implemented, for Italian, in the lexical database ItalWordNet (IWN) (Roventini et al., 2003).

2 The source lexical resource

Semantic relations play a prominent role in the SIMPLE model and this importance explains the marked attention paid to boosting the relational network in a SIMPLE-based lexicon.

The SIMPLE lexical semantic model, whose theoretical framework is the Generative Lexicon (GL) Theory (Pustejovsky, 1995, 2001, 2006), characterizes lexical units through three different descriptive means, namely an ontology and a large set of semantic features and semantic relations which allow to express a wide typology of information.

Entries of the PSC lexicon are in fact semantically classified in terms of the concepts of the SIMPLE ontology, which consists of 157 semantic types organized along hierarchical and non-hierarchical conceptual relations, according to the principle of orthogonal inheritance (Pustejovsky and Boguraev, 1993). Each semantic type is associated to a *template*, which is a schematic structure gathering together the defining properties of the type and imposing therefore well-formedness constraints to the lexical entries candidate to membership.

As to semantic features, they express a wide range of information types, such as the domain of use of the lexical item; properties ─ e.g. *collective, part, edible*, etc. ─ cutting across the type hierarchy and thus allowing to cluster word senses whatever their ontological classification; and traits coherent with the semantic interpretation of *qualia* roles but inexpressible in terms of relations between lexical units.

Semantic relations, as previously stressed, carry great weight in the SIMPLE model. They hold between word senses and are expressed as triplets: <*source semantic unit*, relation, *target semantic unit*>. Their bulk and core is the outcome of a revisitation of Pustejovsky's *Qualia Structure* which, in turn, is inspired from Moravcsik's (1975) interpretation of the Aristotelian modes of explanation (*Aitia*). In GL theory, Qualia Structure is one of the four levels of semantic representation in a generative lexicon[2]. It is composed of four roles (*formal, constitutive, agentive* and *telic*) that specify the multifaceted nature of a word's meaning (*qualia*) and together give "the relational force of a lexical item" (Pustejovsky, 1995 : 76). Qualia roles model the componential aspect of a word's meaning. They enable to connect a word sense to events or entities closely related to its meaning and to capture their role in the lexical semantics of the described word.

In designing the SIMPLE model, the expressiveness of this structure was enriched and gave rise to the *Extended Qualia Structure* (henceforth, *EQS*), wherein each of the four qualia roles subsumes an independent hierarchy of more specific links expressed in terms of semantic relations (60, in total) operating within or across PoS, and which are consistent with the interpretation of their supertype.

Albeit comprehensive and rich, the SIMPLE model has made no provision for a full-fledged descriptive vocabulary allowing to adequately capture the conceptual links holding between events and their participants and among co-participants in events. It provides, however, a fine-grained account of the relations linking animates to their inherent activities.

Encoding in a lexical database the whole set of relationships linking both events to their

[1] The SIMPLE model has been the main source of inspiration for the ISO standard for NLP lexicons, namely the *Lexical Markup Framework*.

[2] The three other levels are: Argument Structure, Event Structure and Lexical Typing Structure.

participants and co-participants in events is most relevant, though. It provides crucial lexical knowledge for enhancing NLP tasks such as information retrieval, information extraction, text understanding, summarization and question answering and is most helpful for Web-based tasks.

Yet, in the IWN lexicon, while synonymic and taxonomic links are largely and consistently encoded, these semantic relations are neither systematically nor extensively implemented, but only sparsely instantiated.

These are the main reasons that induced us to endow *Simple_PLUS* with such information.

From the operational point of view then, borrowing the description vocabulary from the EWN model rather than creating new SIMPLE relations to fill this gap in expressive means was deemed most reasonable. Firstly, because EWN and SIMPLE models present many compatible aspects despite a few differences in some important respects (Ruimy, 2006). Secondly, because this move is in line with the trend, fostered by the international scientific community, aiming at the standardization of lexical description, the interoperability of electronic language resources and the interchange of their content.

3 The Simple_PLUS lexicon

Presently, the semantic lexicon *Simple_PLUS* consists of 30,000 entries, out of which 3,525 adjectives, 20,900 nouns and 5,575 verbs, projected over the whole set of semantic types of the ontology.

Besides its semantic typing, a *Simple_PLUS* entry encodes a vast number of information among which a short definition, an illustrative example, the domain of use of the word sense, a set of distinctive semantic features and semantic relations. Each predicative word sense is moreover related to a set of specific information, namely the type of event it denotes and its predicative representation which consists of its relationship to the predicate[3], as well as the predicate-argument structure with the description of the arguments in terms of semantic role and selectional preferences.

Relational information is at the heart of *Simple_PLUS*. In order to represent the relationships holding among the 30,000 word senses, 73,650 semantic links were encoded, which were expressed by means of 133 different types of semantic relations. The whole apparatus of *Simple_PLUS* lexical semantic relations consists of i) SIMPLE relations, and namely Extended Qualia relations; synonymic, derivational and logical polysemy relations as well as ii) EWN-borrowed relations.

3.1 Extended Qualia Relations

In the *EQS*, the variety of relations interpreting each qualia role allows to gain insight into the relationships holding between word senses, on both the paradigmatic and syntagmatic axes. Not only do they permit to express that an entity has a function, an origin and a composition, they also enable to specify the *type* of its internal constitution, origin and functionality.

In the formal role, which distinguishes an entity within a larger domain, the basic lexical relation 'isa' accounts for taxonomic or troponymic links in the semantic organization of nouns and verbs. By contrast, the relation structuring the adjective class is the antonymic one[4], just as in WordNet and EuroWordNet (Fellbaum et al., 1983: 27). Three subtypes of antonyms are distinguished: complementary antonyms (*alive/dead*) (Cruse 1986: 198-201); gradable antonyms (*cold*) (Cruse, 1986: 204-206), and multiple oppositions (*Italian / French / English*) (Bartning 1980: 112-113).

In the constitutive role, EQ relations provide many different ways to characterize an entity's constitution. They allow to express its composition 'is_made_of' (*mina*, *grafite*) [lead, graphite]; they distinguish between two different interpretations of meronymic (and corresponding holonymic) links, i.e. constituent part 'is_a_part_of' (*capitolo*, *libro*) [chapter, book], membership 'is_a_member_of' (*ministro, governo*) [minister, government] or relationship 'kinship' (*puledro, cavallo*) [foal, horse]. They provide means to characterize different types of intrinsic properties, e.g.: for humans, 'has_as_property' (*tirchio, tirchieria*) [stingy, stinginess], 'uses' (*violinista, violino*) [violinist, violin]; for animals, 'constitutive_activity' (*serpente, strisciare*) [snake, to crawl]; for other entities 'has_as_colour' (*limone, giallo*) [lemon, yellow], 'produces' (*arancio, arancia*) [orange tree, orange]. They also characterize entities

[3] Through links such as: *agent / patient / event_nominalization*.

[4] Only a few taxonomies (such as 'colours', for example) can be isolated within the class of adjectives.

with respect to their locations: 'is_in' and 'lives_in', for geopolitical areas; 'typical_location', encoding natural or artifactual shelters for animals.

In the agentive quale, two main subtypes of *EQ* relations allow to characterize differently the origin of natural entities and events from the one of artifactual entities. The former class includes relations such as 'agentive' and 'agentive_prog', which relates a human to the past or ongoing event he is named after (*fondatore, fondare*) [founder, to found], (*viaggiatore, viaggiare*) [traveller, to travel]; 'agentive_experience' (*sensazione, provare*) [sensation, to feel] or 'caused_by' (*vaiolo, virus*) [smallpox, virus], while in the class of artifactual entities, those properly created are discriminated from those derived from pre-existing ones, e.g.: 'created_by' (*litografia, stampare*) [lithograph, to print] vs. 'derived_from' (*carta, cellulosa*) [paper, cellulose].

In the telic role, different subtypes of *EQ* relations enable to express in different ways the function or purpose of an entity. In the 'instrumental telic' subtype, the relations 'used_for' (*quaderno, scrivere*) [copybook, write]; 'used_as' (*metano, combustibile*) [methane, combustible]; 'used_against' (*antidoto, veleno*) [antidote, poison] and 'used_by' (*telescopio, astronomo*) [telescope, astronomer] characterize the user and the different perspectives according to which the use of an artefact or a substance is perceived. On the other hand, the 'direct telic' relation 'object_of_the_activity' links an entity to the characteristic activity it is the instrument of, i.e. (*aereo, pilotare*) [plane, pilot]. In the 'activity' subtype, three different relations link humans to their activities 'is_the_activity_of' (*insegnante, insegnare*) [teacher, teach], abilities 'is_the_ability_of' (*arrampicatore, arrampicare*) [rock-climber, climb] or habits 'is_the_habit_of' (*fumatore, fumare*) [smoker, smoke].

3.2 New relation set: semiautomatic instantiation

In this section, we focus on the descriptive vocabulary that was imported from the EWN model and the implementation of these new relations in *Simple_PLUS*.

The enrichment of the lexical relation set can only be deemed beneficial if the process is inexpensive in terms of time and effort. Accordingly, steps were taken beforehand to evaluate the feasibility of this enhancement and to provide an estimate of the potential effort needed to achieve the results. Then, the most salient elements of these EWN descriptive means[5] were selected and imported and the tests providing the criteria governing the appropriateness of the relations (Alonge, 1996: 32-34; Climent et al., 1996: 48) were applied. Next, strategies were designed for i) eliciting candidate entries to be paired, through appropriate queries on explicit and implicit information existing in the source lexicon and ii) automating to a large extent the encoding of such relationships and the tuning of the existing ones.

3.2.1 Linking events to their participants

In the EWN model, the 'involvement' relation type links static or dynamic situations (2ndOrderEntities) to concrete or abstract entities (1rst or 3rdOrderEntities, respectively) "whose meaning is 'incorporated' in, or connected with, the meaning of the verb itself" (Alonge, 1996: 31). Different subtypes of this relation relate word senses denoting agents, patients, instruments, location and direction to the events they participate in, to some extent.

For the event-agent relation, all candidate members were automatically identified by investigating the argument structure of the event-denoting word and by inverting the terms of some existing telic, agentive and constitutive relations.

Conversely, the automatic elicitation of the patient of an event turned out to be far more complex. So far, the investigation of the argument's roles gave poor results and a different search strategy restricting the range of the candidate nouns by exploiting also the semantic constraints on verb arguments is now being tested.

Note that, in the above two relations, 'agent' and 'patient' are to be construed according to the definition of these Proto-Roles proposed by Dowty (1991: 572), and which was adopted in the framework of the SIMPLE model as well.

Through the 'involved_location' relation, events such as *abitare, curare, insegnare, pregare, vendere*, etc. [to live, treat, teach, pray, sell] were linked to the typical location they take place in. The candidate word pairs were automatically identified by inverting the terms of word pairs linked by the telic relation

[5] For the time being, we did not consider the 'direction' relation and its two subtypes.

'used_for' and whose first member belonged to the hierarchy of Location types.

In the source PSC lexicon, the SIMPLE constitutive relation 'instrument' was used to link events to the typical means ─ construed in the broadest sense of the term ─ involved in their achievement. Acquiring the EWN 'involved_instrument' relation made it possible to differentiate between the typologies of means. Action-denoting verbs were thus linked to typical, concrete means (instruments, vehicles, substances) via the new relation, e.g.: 'involved_instrument' *volare, aeromobile* [to fly, aircraft] whereas the SIMPLE 'instrument' relation was restricted to more generic means, such as body part, e.g.: 'instrument' (*vedere, occhio*) [to see, eye].

In the EWN model, resultative verbs are linked to 1rst or 3rdOrderEntities through the 'involved_result' relation, e.g.: (*ghiacciare, ghiaccio*) [to freeze, ice]. Similarly, in *Simple_PLUS* this relation links resultative predicates to abstract and concrete entities, and the use of the SIMPLE relation 'resulting_state' is therefore restricted to target events, and namely to the resulting state of a transition or a caused event.

3.2.2 The role of entities in events

The EWN 'role' relation and its subtypes are the corresponding converse of the above five 'involved' relations; they link 1rst or 3rdOrderEntities (concrete or abstract nouns) to 2ndOrderEntities (verbs or event denoting nouns)

Concerning the 'role_agent' relation, the corresponding information is indeed represented in the SIMPLE model, and even more fine-grainedly. Different relations are in fact distinguished, which express the link between an agent entity and an event, according to the semantic type the agent-denoting word belongs to. In the sub-hierarchy of the type Human, the relations are the telic ones 'is_the_activity_of' (*venditore, vendere*) [seller, to sell] for the type Profession; 'is_the_ability_of' (*pittore, dipingere*) [painter, to paint] and 'is_the_habit_of' (*fumatore, fumare*) [smoker, to smoke] for Agent_of_Persistent_Activity; and the agentive ones 'agentive' (*assassino, uccidere*) [murderer, to murder] and 'agentive_prog' (*viaggiatore, viaggiare*) [traveller, to travel] for Agent_of_Temporary_Activity. For the Animal type hierarchy, the link between an agent and the event it is involved in is expressed by the constitutive relation 'constitutive_activity' (*uccello, volare*) [bird, to fly].

Consequently, the 'role_agent' relation was not implemented in *Simple_PLUS* and our attention was rather devoted to checking and incrementing, when needed, the instantiation of the above-mentioned six SIMPLE corresponding relations.

As already observed in section 3.2.1, the links holding between patient entities and the events they participate in are the least easy to automatically derive; therefore, only a few instances of 'role_patient' links were encoded so far.

The relation between an instrument and the action it is used for, and between a location and the typical event that takes place in it were too loosely expressed, in the SIMPLE lexicon, by means of one and the same relation, namely 'used_for' (*pistola, sparare*) [gun, to shoot], (*chiesa, pregare*) [church, to pray]. The acquisition of the two expressive means 'role_instrument' and 'role_location' has allowed to discriminate between the two types of functions. The shift of the previously encoded word pairs to one or the other of these two acquired relations was automatically determined by the type membership of the first term (*pistola*: Instrument; *chiesa*: Building). The newly encoded pairs were then checked against their corresponding 'involved' ones.

3.2.3 Relating co-participants in events

The third type of EWN relations taken into consideration, namely 'co-role' relations, link together co-participants in an event. Therefore, they may relate concrete and abstract entities but not entities and events. So, while 'involvement' and 'role' relations are "type-shifting" relations which operate across parts of speech, 'co-roles' relations are termed "partially type-persistent" relations (Vossen, 2002 : 31).

Six different subtypes of 'co-role' relations are provided for in the EWN model. They link i) agents to patients, to instruments and to results; ii) patients to instruments and to results and iii) instruments to results. Each of these relations has a corresponding converse one. The sole of these conceptual relations expressible in SIMPLE parlance is the one linking an agent and its typical instrument 'uses' (*sarto, ago*) [tailor, needle], and the converse one 'used_by' (*bisturi, chirurgo*) [lancet, surgeon]. These two SIMPLE

relations were therefore maintained in *Simple_PLUS* and the relations types i) and iii) were imported from EWN. Relation types ii) will be taken into consideration once the strategy for searching for patients of events is perfected.

Only a very small number of 'co-role' links were instantiated so far, e.g.: 'co_agent_patient' (*medico, paziente*) [doctor, patient][6]; 'co_agent_result' (*panettiere, pane*) [baker, bread]; 'co_instrument_result' (*fotocopiatrice, fotocopia*) [photocopier, photocopy] and, obviously, an equal number of converse corresponding relations. Yet, incrementing this small set is merely a question of time since the automatic elicitation of word pairs candidate to fill these relations does not pose any particular problem. In fact, relations such as 'co_agent_result' and 'co_result_agent' can be automatically encoded for deverbal result nouns by exploiting the 'involved_result' relation, e.g.: (*costruire, costruzione*) [to build, building] and substituting the first term of the relation for the semantic unit encoded as 'agent nominalization' of the predicate, hence (*costruttore, costruzione*) [builder, building].

3.3 The enrichment process: first evaluation

The enrichment of *Simple_PLUS* with these new relations turned out a largely automated, inexpensive process, mainly based on the reuse and manipulation of existing data for the induction of new information.

The workload was undoubtedly lightened and speeded up thanks to the quality of the source lexical resource. The remarkable wealth of semantic information, the data consistency ensured by the template-driven encoding methodology, the possibility provided by the lexicon management tool to inquire into every single syntactic and semantic property through a tangle of queries and constraints (Ruimy and Toral, 2008) have all facilitated the identification and extraction of candidate entries to be paired, thus making the task worth performing.

The retrieved candidate pairs were then submitted to manual inspection before the new relations were instantiated, in order to prune possible errors imputable to a misencoding in the source lexicon.

Finally, routines were run, which allowed automating, to a large extent, both the insertion of the additional links and the modification of the existing ones.

Identifying these new links through the exploitation of existing data has moreover yielded the non negligible side benefit of permitting an overall consistency check of the lexical resource, since it implied revising and tuning, if necessary, the semantic description.

New relations	Nb. instantiations
involved_agent	1804
involved_patient	26
involved_location	516
involved_instrument	1135
involved_result	82
role_patient	23
role_instrument	1064
role_location	516
co_agent_patient	8
co_patient_agent	8
co_agent_result	13
co_result_agent	13
co_instrument_result	6
co_result_instrument	6

Table 1: Instantiations of new relations

Out of the 5,220 instantiations of the relations borrowed from the EWN model, more than 50% are fully integrable in the Extended Qualia Structure: 1,580 instances are in fact quite harmonized with SIMPLE telic relations while 1,217 are perfectly compatible with the constitutive ones[7]. Since *EQS* is a flexible structure wherein revision and extension processes do not alter the setup insofar as the integrations are consistent with the different roles, introducing these new relations types in the appropriate *EQS* hierarchies was a straightforward task.

The integration in the lexicon of the above-described new links, which come in addition to the rich set of lexical semantic relations belonging to the SIMPLE model, has given a new impulse to *Simple_PLUS* semantic description. In fact, the whole network of relations encodes now essential knowledge for interpreting situation types.

[6] This problem will be overcome when the 'involved_patient' and 'role_patient' relations are populated.

[7] Respectively, 'role_instrument / location' (↔ telic) and 'involved_instrument / result' (↔ constitutive)

For a given situation, information is provided about the type of event, the participating actors, their relationship to the event, their reciprocal links, the instrumental means at stake (together with their components, creation mode and function), the possible outcomes of the event, the links between instrumental means and outcomes, those between means and actors and the spatial location of the event. Not to mention the physical, psychological, spatial and temporal properties encoded in the adjectival entries.

id = "USemD2896anestetizzare"
(to anaesthetize)
Example = "l'anestesista ha anestetizzato il paziente prima dell'intervento chirurgico"
Free_definition = "sottoporre qualcuno ad anestesia"
Semantic_features= "TemplateCauseChangeofState TemplateSuperTypeCauseRelationalChange EventTypeTransition Domain_SURGERY Domain_ANESTHESIOLOGY"
Relations:
***semrel**[8] **= "Isa"**
target = "USemD5415cambiare" (Cause_Change)
***semr = "SRAgentiveCause"**
target = "USem79947fare" (Cause)
***semr = "SRResultingState"**
target = "USem75820addormentato" (PhysicalProp)
****semrel = "InvolvedAgent"**
target = "USem1984anestesista" (Profession)
****semrel = "InvolvedPatient"**
target = "USem80180operando" (Patient_of_Event)
****semrel = "InvolvedLocation"**
target = " USemD79213sala_operatoria" (Building)
****semrel = "InvolvedInstrument"**
target = "USem3278siringa" (Instrument)
****semrel = "InvolvedInstrument"**
target = "USem3036anestetico" (Substance)
Predicative Representation:
Predicate = "PREDanestetizzare#1"
typeoflink = "Master"
Arguments:
id = "ARG0anestetizzare#1"
semanticrole = "Role_ProtoAgent"
select_restr= "PLUS_ HUMAN"
id = "ARG1anestetizzare#1"
semanticrole="Role_ProtoPatient"
select_restr ="PLUS_ ANIMATE"
id = "ARG2anestetizzare#1"
semanticrole="Role_Underspecified"
comment="Shadow_argument"
select_restr ="USem3036anestetico"

Table 2: A semantic entry: *to anaesthetize*

[8] *=SIMPLE relations; **= EWN-borrowed rels.

3.4 Optimizing the lexicon format

The database management tool of the PSC lexicon, which is the one used for *Simple_PLUS,* does not allow for the computation of inheritance at the semantic level. Consequently, although many properties are largely shared and could therefore be inherited from their ancestors' entries, every single feature of a semantic unit has to be explicitly defined in its lexical entry.

Undoubtedly, the addition of more than 5,000 relations has still exponentially increased redundancy. Some entries[9] contain so much information that they turn unmanageable unless an inheritance mechanism enters the picture so as to permit to overtly represent only those word's specific properties and links that are essential to discriminate it from its closest semantically related words, especially its hyperonym. This presupposes, of course, to rely on a high-quality encoding and particularly on consistent taxonomic links.

The implementation of inheritance, which is currently being tested (Del Gratta et al., 2008), is providing encouraging results, i.e. a dramatic reduction of explicitly encoded links. To give but one example, the lexical entry for the main meaning of the verb *vendere* [to sell], which is involved (as source or target term) in 273 semantic relations is reduced by 250 links that are derived by inheritance whereas only 23 specific relations are overtly represented.

4 Concluding remarks

The extensive instantiation in *Simple_PLUS* of new relations linking both events to their participants and co-participants in events helps gain deeper knowledge of the syntactic and semantic behaviour of word senses. It strengthens and enhances the representation of the semantic predicate.

[9] such as, for example, those of high frequency activity verbs, as for example 'to work', encoding the link to their typical agents.

On the one hand, while in the SIMPLE model predicate's arguments are constrained through restrictions on their semantic type membership, the newly encoded links enable to move forward from the expression of combinatorial possibilities at the ontological level to their specification at the lexical level.

On the other hand, the relations involving instruments, locations and results enrich the semantic description by providing knowledge on those adjuncts or extra-thematic roles which are part of a semantic scenario and are therefore essential for a full understanding of texts.

Combined with the wealth of relations provided by the SIMPLE model, the newly encoded links between events and entities constitute powerful tools that contribute to performance gains in NLP applications and are most relevant to build up and make explicit the semantic scenarios potentially useful to the Semantic Web.

References

Alonge, A. ed. 1996. Definition of the links and subsets for verbs, EuroWordNet Project LE2-4003, Deliverable D006, WP4.1, final version.

Climent, S., H. Rodríguez and J. Gonzalo. 1996. Definition of the Links and Subsets for Nouns of the EuroWordNet Project. Deliverable D005, EuroWordNet, LE2-4003, Computer Centrum Letteren, University of Amsterdam.

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

Del Gratta, R., N. Ruimy and A. Toral. 2008. SIMPLE-CLIPS ongoing research: more information with less data by implementing inheritance. In *Proceedings of LREC 2008*, CD-ROM, Marrakech (Morocco).

Dowty, D. 1991. Thematic Proto-roles and Argument Selection. *Language*, 67(3): 547-619.

Fellbaum, C., D. Gross and K. Miller. 1993. Adjectives in WordNet, in Miller et al. *Five Papers in WordNet*, pp. 26–39, Cognitive Science Laboratory, Princeton University.

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, 404 pp.

Moravcsik, J. M. 1975. Aitia as Generative Factor in Aristotle's Philosophy, *Dialogue* 14 : 622-36.

Pustejovsky, J. and B. Boguraev. 1993. Lexical Knowledge Representation and Natural Language Processing, *Artificial Intelligence* 63 : 193-223.

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

Pustejovsky, J. 2001. Type Construction and the Logic of Concepts. In P. Bouillon and F. Busa eds., *The Syntax of Word Meanings*, Cambridge University Press, pp. 91-123.

Pustejovsky, J. 2006. "Type Theory and Lexical Decomposition". In *Journal of Cognitive Science* 6: 39-76.

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, XVIII-XIX, Tomo II, pp. 745-791, IEPI. Pisa-Roma.

Ruimy, N., M. Monachini, R. Distante, E. Guazzini, S. Molino, M. Ulivieri, N. Calzolari, and A. Zampolli. 2002. CLIPS, A Multil-level Italian Computational Lexicon: a Glimpse to Data. In *Proceedings of LREC 2002*, Volume III, pp. 792-79, Las Palmas de Gran Canaria (Spain).

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, XVIII-XIX, Tomo II, pp. 821-864, IEPI. Pisa-Roma.

Ruimy, N. 2006. Merging two Ontology-based Lexical Resources. In *Proceedings of LREC 2006*, pp. 1716-1721, Genova (Italy).

Ruimy, N. 2007. Enhancing SIMPLE Semantic Relations: A Proposal**.** In *Proceedings of 3rd Language & Technology Conference*. pp. 119-123, Fundacja Uniwersytetu im A. Mickiewicza, Poznań (Poland).

Ruimy, N. and A. Toral. 2008. More semantic links in the SIMPLE-CLIPS database. In *Proceedings of LREC 2008*, CD-ROM, Marrakech (Morocco)**.**

Vossen, P. ed. 2002. EuroWordNet General Document, final version, Vrije Universiteit, Amsterdam.