Merging two Ontology-based Lexical Resources

Nilda Ruimy
Istituto di Linguistica Computazionale, Consiglio Nazionale delle Ricerche
Via G. Moruzzi 1- 56124 Pisa, Italy
nilda.ruimy@ilc.cnr.it

Abstract

ItalWordNet (IWN) and PAROLE/SIMPLE/CLIPS (PSC), the two largest electronic, general-purpose lexical resources of Italian language present many compatible aspects although they are based on two different lexical models having their own underlying principles and peculiarities. Such compatibility prompted us to study the feasibility of semi-automatically linking and eventually merging the two lexicons. To this purpose, the mapping of the ontologies on which basis both lexicons are structured was performed and the sets of semantic relations enabling to relate lexical units were compared. An overview of this preliminary phase is provided in this paper. The linking methodology and related problematic issues are described. Beyond the advantage for the end user to dispose of a more exhaustive and in-depth lexical information combining the potentialities and most outstanding features offered by the two lexical models, resulting benefits and enhancements for the two resources are illustrated that definitely legitimize the soundness of this linking and merging initiative.

1. Introduction

The two largest electronic, general-purpose lexical resources of Italian language were developed during the last decade at the National Research Council Institute for Computational Linguistics (ILC), in Pisa.

ItalWordNet[1] (henceforth IWN), a lexical semantic database, was created in the framework of the *EuroWordNet* (EWN) project[2] (Vossen, 1999) and then extended during the Italian national project *Sistema Integrato per il Trattamento Automatico della Lingua*[3] (SI-TAL), which was coordinated by A. Zampolli and devoted to the creation of large linguistic resources and software tools for processing Italian written and spoken language. In the SI-TAL IWN database, the principles and linguistic design of the EWN model were fully maintained. The only aspects in which IWN differs are a few amendments made to the ontology in order to allow for the representation of adjectives and the identification and addition of further lexical-semantic relations.

PAROLE/SIMPLE/CLIPS[4] (henceforth PSC[5]), a multi-layered lexicon, was developed over three different projects. The morphological and syntactic models and the kernel of related lexicons were elaborated in the EU *LE-PAROLE* project; the semantic model and the core of the semantic lexicon, in the EU *LE-SIMPLE* project[6]; the phonological level of description and the extension of the lexical coverage were, on the other hand, performed in the context of the Italian national project *Corpora e Lessici dell'Italiano Parlato e Scritto*[7] (CLIPS).

Since IWN is a one-layer lexical database, the only comparable information is the semantic one. In this regard, the EWN and SIMPLE lexical models do have different underlying principles and peculiarities such as, for example, a different ontological framework and a different organization of lexical units. However, IWN and PSC lexicons also present many compatible aspects[8] that prompted us to envisage the semi-automatic link of the two lexical databases, eventually merging the whole information into a common representation framework. Roventini *et al.* (2002) expressed relevant preliminary considerations regarding the linking, as the result of a feasibility study. Since then, a separate use of IWN and PSC lexical data confirmed us in our conviction that linking the two databases is a most appropriate and timely initiative. A linking methodology was therefore developed whose implementation in an Access software tool started in early 2006, at ILC.

Besides the advantage for the end user to dispose of a more exhaustive and in-depth lexical information combining the potentialities and most outstanding features of the two lexical models, the linking process also lets inconsistencies that unavoidably exist in both resources emerge, allowing therefore to amend them. It also makes it possible to enrich each resource by relevant information types that are peculiar to the other's model.

2. ItalWordNet

The IWN lexical semantic database (Roventini *et al.*, 2003) contains the semantic description of 67,000 word senses of verbs, common and proper nouns, adjectives, adverbs and multi-word units which are clustered in about 50,000 *synsets* (i.e. synonym sets). Synsets, that represent lexicalized concepts, are classified in terms of the Top Ontology and interconnected by means of a set of semantic relations. In EWN, about 1,000 *Base Concepts* representing the most important top nodes of hyp(er)onymic hierarchies were linked to one or more concepts of the Top Ontology. The linking process went on in IWN, taking into account further top nodes and related hierarchies, e.g. the adjectives. One of the salient features of the resource is the connection of all IWN synsets to the Princeton Wordnet database (Miller *et al.,* 1990).

[1] http://www.ilc.cnr.it/viewpage.php/sez=ricerca/id=834/vers=ing
[2] A project of the EC Language Engineering programme (LE4003) http://www.hum/uva.nl/~ewn.
[3] 'Integrated System for the Automatic Treatment of Language'.
[4] http://www.ilc.cnr.it/clips/CLIPS_ENGLISH.htm
[5] PSC is not the acronym of the lexicon. It is only used here for the sake of brevity.
[6] http://www.ub.es/gilcub/SIMPLE/simple.html
[7] 'Corpora and Lexicons of spoken and written Italian'
[8] It is worth reminding, in this regard, that EWN was one of the inspiration sources for the SIMPLE model of semantic representation.

2.1. IWN Top Ontology

The IWN Top Ontology (henceforth, TO) underwent a few changes wrt. EWN Top Ontology in order to account for the semantics of new lexical categories. It is a hierarchy of 65 language-independent Top Concepts (henceforth, TCs) that express fundamental semantic distinctions[9]. TCs are clustered in three main categories distinguishing three types of entities: 1stOrderEntity, 2ndOrderEntity and 3rdOrderEntity. The subclasses of these categories are hierarchically ordered by means of a subsumption relation and also structured in terms of (disjunctive and non-disjunctive) opposition relations. IWN TCs are inherited through the link to an Interlingual-Index (ILI), which is an unstructured version of WordNet 1.5.

2.2. Semantic Representation

In IWN, a word sense is defined by the ontological classification of the synset it belongs to and also by means of a rich set of lexical-semantic relations that link intracategorial and intercategorial synsets (Alonge *et al.* 1998). The most relevant relations encoded are hyponymy, synonymy, antonymy, meronymy and xpos_near_synonymy. The last one, which is a cross-categorial relation, was defined in EWN in order to link the different lexicalizations of a concept, which is of utmost importance for IR applications.

3. PAROLE-SIMPLE-CLIPS

PSC (Ruimy *et al.*, 2002), a general purpose lexicon, consists of 55,000 one-word lemmas (verbs, nouns, adjectives, adverbs and grammatical words) with phonological, morphological and syntactic description and 55,000 word senses encoded at the semantic level[10], all in full accordance with the international standards set out in the PAROLE-SIMPLE model (Ruimy *et al.*, 1998; Lenci et al., 2000). Along the description levels, a PSC entry defines the whole set of phonological, morphological and inherent syntactic and semantic properties of a headword. Its syntactic subcategorization frame(s) is (or are) described in terms of optionality, syntactic function, syntagmatic realization as well as morpho-syntactic, syntactic and lexical properties of each slot filler. As far as semantics is concerned, the theoretical approach adopted by the SIMPLE model is essentially grounded on a revisited version of some fundamental aspects of the Generative Lexicon (GL). The semantic lexicon is structured in terms of the SIMPLE ontology of semantic types and the basic unit which is described is a word sense, represented by a *semantic unit*.

[9] It is worth noting that TCs are to be construed more as lexical features than as real conceptual classes.

[10] Part of the syntactic and semantic encoding was performed by the THAMUS Italian Consortium for Multilingual Documentary Engineering.

3.1. SIMPLE Ontology

The SIMPLE Ontology (SO) consists of 157 semantic types designed for the multilingual lexical encoding of concrete and abstract entities, properties and events. It is a multidimensional type system, based on both hierarchical and non-hierarchical conceptual relations and which takes into account the principle of orthogonal inheritance (Pustejovsky & Boguraev, 1993). The SO reflects the GL assumption that lexical items are multidimensional entities which present various degrees of internal complexity and thus call for a lexical semantic description able to account for different ranges of meaning components. Accordingly, the SO consists of *simple* types — which can be fully characterized in terms of one-dimensional taxonomic relations — and *unified* types — which also incorporate orthogonal meaning dimensions[11] and thus require a multidimensional organization. According to the philosophy governing the SIMPLE ontology, assigning a semantic type to a lexical unit does not simply mean ascribing it a mere semantic label but rather endowing it with a structured set of semantic information that is crucial to the type's definition[12].

3.2. Semantic Representation

A SIMPLE-based semantic representation consists primarily in a semantic type assignment. The membership of a word sense in an ontological type then inherently triggers the instantiation of a rich bundle of semantic features and relations. Among these are the 60 relations of the *Extended Qualia structure*, a revisited version of the original GL representational tool that enables to describe both the componential aspect of a word meaning and its relationships to other lexical items. Predicative word senses are also characterized as to their contextual information, expressed by a semantic predicate whose arguments are described in terms of thematic roles and semantic constraints. Lexical information is interrelated across the different description levels. Between syntax and semantics then, the link occurs through the projection of the predicate-argument structure to its syntactic realization(s).

4. The Ontological Frameworks

As observed, in both lexicons, an ontology of semantic types provides the backbone for knowledge organization. Semantic information is also expressed by means of semantic features and relations. A primary phase in the process of integrating the two lexicons consists therefore in the comparison of their ontological framework.

[11] This multidimensionality is expressed in SIMPLE ontology by means of the *Extended Qualia Structure* which played a crucial role in defining the distinctive properties and differentiating the degrees of complexity of the semantic types.

[12] The definitorial properties of each type are gathered in a schematic, underspecified structure called *template* that supplies the lexicographer with clusters of structured, language-independent information specific to a semantic type. This template-driven encoding methodology ensures a high level of descriptive completeness and consistency.

A first, general observation is that IWN TO consists of a set of rather flat top semantic features whereas SO encompasses mono- and multi-dimensional types with associated templates of structured information that define the content of the conceptual types. Before comparing the ontological classes, let us very briefly illustrate the ontological typing in the two models.
According to the SIMPLE model, a semantic unit is associated to one single semantic type. The EWN/IWN model, by contrast, allows for a multiclassification: synsets are seldom linked to one single ontological node but rather cross-classified in terms of multiple, non disjoint, TCs[13], as illustrated in table 1. Noteworthy here is that, in IWN, the ontological classification is determined by the choice of the synset hyperonym.

w.sense	SIMPLE semantic type	IWN Top Concepts
tavolo table	Furniture	Artifact Object Furniture
metro meter	Unit_of_measurement	Purpose, Quantity, Social, Static, Usage
marxismo marxism	Movement_of_thought	3rdOrder Mental
spostare to move	Cause_change_location	Cause, Dynamic, Location

Table 1: PSC vs. IWN ontological typing

Although formally different, the information provided by these two types of ontological classification is substantially equivalent: thanks to the multidimensional nature of the ontology, SIMPLE types encompass in fact the various meaning dimensions that are expressed in IWN by the different TCs cross-classifying 1st and 2ndOrder Entities.

4.1. Ontology Mapping

By ontology mapping process, we intend here a schema mapping level, i.e. the establishment of correspondences between the conceptual classes of both ontologies with a view to further matching their respective instances. A manual mapping[14] from IWN to SIMPLE ontology for entities and events[15] was performed, which will be briefly outlined in the following.
The first subdivision level of IWN TO consists of three main classes:
The 1stOrderEntity class structures concrete entities (referred to by concrete nouns). Its main cross-classifying subclasses: Form, Origin, Composition and Function correspond to the four Qualia roles the SIMPLE model avails of to express orthogonal aspects of word meaning. Their respective subdivisions consist of (mainly) disjoint classes, e.g. Natural vs. Artifact, Substance vs. Object. To each one corresponds, in most of the cases, a SIMPLE semantic type (correspondence 1,1) or a type hierarchy (1,n) belonging to the Concrete_entity top type. Some other TCs, such as Comestible, Liquid, are instead mappable to SIMPLE distinctive features: Plus_Edible, Plus_Liquid, etc.
The 2ndOrderEntity class classifies static or dynamic situations denoted by nouns and verbs, adjectives and adverbs. 2ndOrderEntities are primarily characterized in terms of two classification parameters: *SituationType* – whose two disjoint features, *static* and *dynamic* encode the event structure – and *SituationComponent,* which subsumes a set of combinatorial classes providing a more conceptual classification in terms of semantic components of a concept, e.g.: *Manner*, *Experience*, *Communication, Cause*.
Concerning the Situation Type, in the SIMPLE model, the event structure is expressed by means of the three-valued feature Eventtype = *state*, *process*, *transition*, values which correspond in IWN respectively to *static*, *unbounded event* and *bounded event*. As to the combinatorial subclasses of the Situation Component, each one generally corresponds to one or more SIMPLE types, depending on the Situation Type value and/or the other Situation Components it combines with, e.g.:

IWN Top Concepts	SIMPLE semantic type
Existence Bounded Cause Physical	Creation, Cause_natural_transition
Existence Static	Exist
Experience Mental Dynamic	Experience_event, Modal_event
Experience Physical Stimulating Dynamic	Perception

Table 2: TCs and semantic type correspondences

The 3rdOrderEntity class, which has no further subdivision classifies abstract entities (denoted by abstract nouns) existing independently of time and space. These entities fall into the Abstract_entity type hierarchy of the SIMPLE ontology.
A preliminary comparison of the ontological typing of adjectives was also performed. However, these entities will be dealt with in a further phase of work.
Notwithstanding the different approaches taken for their design and some different underlying principles, these two ontologies globally show a significant degree of overlapping and no fundamental difference in conceptualization. Three general remarks are in order here:

- The mapping between the two type systems, which has been carried out on a theoretical basis, presupposing correspondences between their respective concepts according to their definition, clearly requires some tuning in order to cope with the actual ontological classification in the two resources.
- Owing to the different extension of both ontologies, some specific concepts – which are expressed in SIMPLE Ontology by lower level semantic types – are likely to have no equivalent in IWN TO.
- Not surprisingly, mapping from event-denoting PSC semantic units to IWN 2ndOrderEntities immediately appears much more challenging than dealing with 1rstOrderEntities that pose less tricky problems.

[13] The more specific the word, the more TCs contributing to its description.
[14] A first and partial version of the mapping was already performed in Ruimy, Gola and Monachini, 1998.
[15] Ruimy (2005) http://www.ilc.cnr.it/clips/Ontology_mapping.doc

Due to space limitations, the comparison which has been performed of the semantic relations used by both models to relate senses along the paradigmatic and syntagmatic axes will not be illustrated in this paper.

5. Linking the Lexical Resources

5.1. Linking Methodology

Owing to a different organizational structure of information in the two resources, the linking process involves elements having a different status, viz. autonomous semantic units (henceforth, SemU) in PSC and synsets clustering 1 to n word senses — referred to as 'variants' — in IWN.
In a preliminary phase, mapping runs are used as a testbed for assessing both the linking methodology and software performance and for tuning or relaxing mapping conditions. In order to focus on these aspects and to avoid dealing with huge, unmanageable sets of data, a semantic type-driven process is adopted in the training phase.

5.1.1. Mapping from PSC to IWN

Taking as starting point a semantic type of the SIMPLE ontology, say 'Instrument', and considering all its SemUs along with their PoS and 'isa' relation, we explore the IWN resource in search of linking candidates whose ontological classification fulfils the matching requirements (see example in table 2).
Two sets of data are returned from each mapping run:

1) Matched pairs of word senses, i.e SemUs and synset variants with identical string, PoS and whose ontological classification perfectly match, according to the Ontology Mapping table.

	SIMPLE semantic types	IWN TopConcepts
1	Instrument	Artifact Object Instrument
2	Cognitive_fact	3rdOrder Mental
3	Cause_change; Cause_relational_ change; Cause_change_of_state	Cause BoundedEvent

Table 2: Fragment of the Ontology Mapping Table

After human validation, these matched word senses are linked. In example 1, linking occurs between a SemU and a *one-variant* synset:

1. SemU2838zappa ↔ synset15564 {N *zappa*1}

In example 2, on the other hand, linking first occurs between a SemU and one word sense of a *multi-variant* synset:

2. SemU60203*migliorare* ↔ synset38725 {V …, *migliorare*2,…}

Consistency then implies that the other variants (viz. *perfezionare*1, *aggiustare*2) map to PSC entries sharing the same semantic type as SemU60203*migliorare*. This is indeed the case here and, as shown below, all synset variants are therefore linked to their corresponding PSC entries.

PSC semantic type = Cause_change:
… SemU60203***migliorare*,V**, …., SemU 3938*modificare*,V…, SemU79575*perfezionare*, V, …., SemU59154*aggiustare*, V,…
↕
IWNsynset38725 : ontological classification = Cause BoundedEvent
{**V** *perfezionare*1, ***migliorare***2, *aggiustare*2}

Such a situation does not, however, systematically occur and cases have been observed whereby, especially as far as event denoting words are concerned, SemUs corresponding to variants of the same synset do not share a common SIMPLE semantic type.

2) Unmatched word senses, as e.g. 'Instrument'-typed SemUs not matching to 'Artifact Object Instrument'-typed IWN synset variants. This list of leftover, unmatched senses contains heterogeneous data. In fact, matching failure may be due either to the mere non-existence, in IWN, of word senses corresponding to PSC entries[16] or to a mismatch of ontological classification between word senses existing in both resources. Focusing on this latter set of data, two main obstacles to matching emerge:
a) an incomplete ontological information:
► As noted in section 4, IWN synsets are cross-classified in terms of a combination of TCs. This combined notation is however sometimes only partially encoded and cases of 1rstOrderEntities lacking some meaning component or 2ndOrderEntities lacking one of the two classifiying parameters are not rare.
For the linking purpose, such a problem may, in a number of cases, be overcome by relaxing the mapping constraints. Yet, this solution can only be applied if the actual ontological classification, in spite of its incompleteness, is still informative. For example, in absence of the complete TC combination reported in equivalence 1 of table 2, mapping can only be licensed if the actual ontological label consists at least of the TC 'Instrument', which is likely to be associated to all instrument-denoting synsets.
Far more problematic to deal with are those cases of incomplete and little informative ontological labels, e.g. 1rstOrderEntities as different as *medicinale, anello, laccio, vetrata* and only classified as 'Function' or 2ndOrderEntities lacking either a Situation Component or a SituationType, e.g. *unirsi* classified as 'BoundedEvent' or *sciogliere* as 'Cause'.
b) a different ontological information:
► Besides mere encoding errors for which a correction phase is foreseen, ontological classification discrepancy may be imputable to:
i) a different but equally defensible meaning interpretation (e.g.: *ala* (aircraft wing) : 'Part' vs. 'Artifact Instrument Object'). Word senses falling into this category are clustered into numerically significant sets according to their semantic typing and then studied with a view to establishing further equivalences between ontological classes or to identify, in their classification schemes, descriptive elements lending themselves to be mapped.
ii) a different level of specificity in the ontological classification, either due to the lexicographer's subjectivity — e.g. instrument-denoting senses classified as 'Artifact' —

[16] A list of unfound lexical units is automatically generated, in the event of a further extension of the IWN lexicon.

or to an objective difference of granularity of the ontologies.
Problems emerging with instances of ii) may be bypassed by climbing up the ontological hierarchy, identifying the parent nodes and allowing them to be taken into account in the mapping process. For example, map 'Cause_change' to 'Cause BoundedEvent' (cf. table 2) but also allow mapping it to 'Cause Dynamic'[17]. Similarly, for PSC word senses belonging to low level semantic types, take into consideration their supertype in order to match them to IWN synsets with a less specific ontological classification.

5.1.2. Mapping from IWN to PSC

The whole process described in 5.1.1.1 is then applied in the reverse sense, viz. IWN synsets variants with a specific ontological classification are mapped onto PSC entries. Besides an identical set of matching lexical units, such process returns the set of IWN lemmas missing in PSC[18] as well as the set of existing but non matching items, due to a totally or partially different ontological typing (cf. 5.1.1.2.a. and b.).

5.1.3. Hyperonymic Links

Hyperonyms of matching candidates are taken into account during the linking process and play a particularly determinant role in the resolution of cases, such as the following ones, whereby matching fails due to a conflict of the ontological classification:
► sets of word senses displaying a different ontological classification in each resource but sharing the same hyperonym, e.g. *collana, braccialetto, orecchino,* etc. typed as 'Clothing' in PSC and as 'Function' in IWN but sharing the hyperonym *gioiello* in both resources.
► polysemous senses belonging to different semantic types in PSC but sharing the same ontological classification in IWN, e.g.: SemU1595*viola* 'Plant' and SemU1596*viola* 'Flower' vs. IWN: *viola*1 (has_hyperonym *pianta*1) and *viola*3 (has_hyperonym *fiore*1), both typed as 'Group Plant'.

6. Enhancement of the Resources

The linking of two resources based on such valuable and widely tested lexical models that have addressed challenging (and complementary) research issues in lexical semantics enables knowledge and information sharing that will translate into reciprocal enhancements for both lexicons.
Entries of the two resources will enrich each other with complementary information types that reflect the different philosophy of the two theoretical models. In EWN, it is the richness of sense distinctions and the variety of semantic relations that are in the foreground. In SIMPLE, on the other hand, focus is put on richly describing the meaning and semantic context of a word and on linking its syntactic and semantic representation, which is crucial for most language processing applications.

[17] The SituationType 'Dynamic', that encodes the event structure, is the parent node of 'BoundedEvent' and 'UnboundedEvent'.
[18] As before, a list of unfound lexical units is generated, in the event of a further extension of the PSC lexicon.

6.1. IWN Information Enriching PSC SemUs

► The organization of lexical knowledge has entailed a quite coherent clustering of synonyms in Wordnet-based resources. The SIMPLE model, on the other hand, has devoted more attention to other relation types and less importance has been given to the instantiation of synonymy links. Integrating the two lexicons, PSC entries could easily be enriched with synonymy links, based on synset membership. In like manner, missing senses of existing words and new lemmas could be quickly and consistently encoded in PSC lexicon.
► In IWN, hierarchical links are of fundamental importance and hence rather consistently expressed by two relations 'has_hyperonym' and 'has_hyponym' that allow a cross-checking of data. In the SIMPLE model, on the other hand, the focus put on covering the whole range of a word's syntactic and semantic uses has sometimes prejudiced the enforcement, in PSC entries, of coherent taxonomic links and yielded cases of circularity. Such cases could be amended by resorting to IWN hyperonymic links for nouns and verbs.
► IWN 'Involved_agent/patient/instrument/location' and 'Role_agent/patient/instrument/location' semantic relations, respectively linking 2nd with 1rst or 3rd OrderEntities and conversely could be most helpful to link more straightforwardly an event to its participants, e.g.: *curare*: involved_agent = *medico*, involved_patient = *malato*, involved_location = *ospedale*; an entity to an event: *alunno* role_patient: *insegnare* or even to relate event's participants to each other: *insegnante* co_agent_patient *alunno.*
These links could moreover allow to discriminate the nature of some relationships that are rather poorly rendered, in the PSC lexicon, by the overused —and hence misused — constitutive relation 'concerns', e.g.: 1) *otturazione* 'concerns' *dente;* 2) *sbarcare* 'concerns' *nave* could be respectively replaced by 'involved_patient' and 'involved_ source_direction'.

6.2. PSC Information Enriching IWN Synsets

► No argument structure information is provided in IWN. Linking the two lexicons, IWN predicative units could be endowed with information concerning their syntactic and semantic subcategorization frames.
► IWN word senses could also inherit the PSC extensively encoded information concerning their domain of use. Such information, most relevant to IR, WSD, IE and parsing, enables – among other – clustering semantically related lexical items pertaining to specific domains, regardless of their PoS and type membership.
► Given the rich lexical information foreseen by the SIMPLE model, IWN synsets could also gain:
- a finer-grained ontological classification: let us observe for example that, as against the 'Plant', 'Human', and 'Communication' TCs, SIMPLE Ontology offers respectively 5, 9 and 8 semantic types, each one providing a rich bundle of specific information;

- a semantic description less prominently based on taxonomic relations. SIMPLE semantic relations, which are defined along multiple dimensions, enable to avoid an overloading of the ‘isa’ relation and to represent senses not correctly definable in terms of the hyperonymic link.
- the expression of further orthogonal meaning dimensions: e.g., synset variants such as *asfalto* or *celluloide*, associated to the TC Substance and bearing constitutive information, could acquire, through their linking to the corresponding PSC entries, agentive and telic dimensions.
- the account of systematic polysemy links. Regular polysemy is expressed, in the SIMPLE model, through distinct entries connected by means of a polysemous relation linking the ontological typing of pairs of senses, according to a set of polysemous classes, e.g. *giornale*: Building / Institution; Location / Human_group. In IWN, such polysemous senses are provided as separate meanings but no mention is made about the way they relate to each other, although the possibility of assigning two or three disjunctive hyperonyms theoretically exists.
- a more specific identification of the nature of some syntagmatic relationships not expressible in the IWN model, and which are, for instance, most relevant for extracting semantic networks, e.g.: *antipiretico* ‘used_against’ *febbre* (antipyretic, temperature), *acetone* ‘used_as’ *solvente* (acetone, solvent).

The linking process will also permit to enhance the consistency of the two resources since it implies a de facto reciprocal assessment of both coverage and accuracy, which is particularly relevant to hand-built lexical resources.

7. Concluding Remarks

Differences regarding the nature of linking units, the granularity of sense distinction and the ontological typing are complex issues to be addressed during the linking process. Problems arise, in particular, when encoding incompleteness or inconsistency generates unpredictable, non-systematic ontological typing discrepancies whereby a theoretical comparison of the models evidences a high degree of overlapping. Nevertheless, the wide range of compatibility the models show induces us to strongly believe that semantic interoperability is indeed achievable. Semantic integration of these resources is all the more desirable considering their multilingual vocation: IWN linked to the WN of seven other languages and PSC sharing with eleven European lexicons a theoretical model, representation language, building methodology and a core of entries.

The linking process has just started with the handling of three main classes of concrete entities. The first results sound quite encouraging since 62% of the word senses considered have been successfully linked as a result of the first mapping run. Unmatched items are now being studied with a view to refining / relaxing the mapping requirements and processing them afresh. In the meanwhile, the mapping process is being carried on with further classes of first-order entities.

References

Alonge, A., Calzolari, N., Vossen, P., Bloksma, L., Castellon, I., Marti, T., Peters, W. (1998). The Linguistic Design of the EuroWordNet Database, Special Issue on EuroWordNet. In N. Ide, D. Greenstein, P. Vossen (eds.), ‘Computers and the Humanities’, XXXII, 2-3, (91--115).

Lenci, A. et al. (2000). SIMPLE Linguistic Specifications, Deliverable D2.1, ILC-CNR, Pisa.

Miller, G. et al. (1990). Introduction to WordNet: An On-line Lexical Database, International Journal of Lexicography, III, 4, 235--244.

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

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

Roventini, A., Alonge, A., Bertagna, F., Calzolari, N., Cancila, J., Girardi, C., Magnini, B., Marinelli, R., Speranza, M., Zampolli, A. (2003). ItalWordNet: Building a Large Semantic Database for the Automatic Treatment of Italian. Computational Linguistics in Pisa, Special Issue, XVIII-XIX, Pisa-Roma, IEPI. Tomo II, 745--791.

Roventini, A., Ulivieri, M., Calzolari, N. (2002). Integrating two semantic lexicons, SIMPLE and ItalWordNet: what can we gain?, LREC 2002, Third International Conference on Language Resources and Evaluation Proceedings, Vol. V, pp. 1473-1477, Las Palmas de Gran Canaria.

Roventini, A., Ruimy N. (2005). Linking and harmonizing different lexical resources: a comparison of verbal entries in ItalWordNet and PAROLE-SIMPLE-CLIPS. In Proceedings of the Third International WordNet Conference, Jeju Island, Korea

Ruimy N., Monachini M., Distante R., Guazzini E., Molino S., Ulivieri M., Calzolari N., Zampolli A. (2002). CLIPS, a Multi-level Italian Computational Lexicon, LREC 2002, Third International Conference on Language Resources and Evaluation Proceedings, Vol. III, (pp.792--799), Las Palmas de Gran Canaria.

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 A. Zampolli, N. Calzolari, L. Cignoni, (eds.), Computational Linguistics in Pisa, Special Issue, XVIII-XIX, (2003). Pisa-Roma, IEPI. Tomo II, 821-864.

Ruimy N., Roventini A. (2005). Towards the Linking of two Electronic Lexical Databases of Italian. In Z. Vetulani (ed.), L&T'05, April 21-23, Poznan, Poland. Wydawnictwo Poznanskie Sp. z o.o. 230-234.

Silva N., Maio P., Rocha J. (2005) An approach to ontology mapping negotiation, Proceedings of the Third International Conference on Knowledge Capture Workshop on Integrating Ontologies; Banff, Canada.

Vossen, P. (ed.) (1999). EuroWordNet General Document. http://www.hum.uva.nl/~EWN.