

Linking and Harmonizing Different Lexical Resources: a Comparison of Verbal Entries in ItalWordNet and PAROLE-SIMPLE-CLIPS

Adriana Roventini

Istituto di Linguistica Computazionale, CNR,
Via G. Moruzzi 1, 56124 Pisa, Italy
adriana.roventini@il.cnr.it and Nilda Ruimy
Istituto di Linguistica Computazionale, CNR,
Via G. Moruzzi 1, 56124 Pisa, Italy
nilda.ruimy@ilc.cnr.it

Abstract

During the last years, in the framework of Computational Linguistics, many lexical resources have been developed which aim at coding complex lexical semantic information according to different linguistic models (WordNet, Frame Semantics, Generative Lexicon, etc.). However, these resources are often not easily accessible nor available in their entirety. Yet, from the point of view of the continuous growth of the technology (Semantic Web), their visibility, availability, integration and harmonization are becoming of utmost importance.

ItalWordNet and PAROLE/SIMPLE/CLIPS are two resources which, tackling lexical semantics from different perspectives and being at least partially complementary could profit from linking each other. In this paper we address the issue of linking these resources focusing on the most problematic part of the lexicon: the second order entities. In particular, after a brief description of the two resources, their different approaches to the verb semantics are described; an accurate comparison of a set of verbal entries is carried out, with a view to evaluating the possibilities and the advantages of a semiautomatic link; finally, the results and the future work are illustrated.

Introduction

ItalWordNet (henceforth IWN) was first developed within the EuroWordNet (EWN)¹ project (Vossen, 1999) and then extended in the framework of an Italian national project for the automatic treatment of the language SI-TAL². IWN (Roventini et al. 2003) is a large lexical-semantic database containing semantic information for about 50,000 synsets of nouns, verbs, adjectives, adverbs, and a subset of proper nouns. The information is encoded in the form of lexical-semantic relations between pairs of synsets (synonym sets).

¹EWN was a project in the EC Language Engineering (LE4003) programme. Complete information on EWN can be found at its web site: <http://www.hum/uva.nl/~ewn>.

²The SI-TAL project: 'Integrated Systems for the Automatic Treatment of Language' was a National Project, coordinated by A. Zampolli, devoted to the creation of large linguistic resources and software tools for the Italian written and spoken language processing. Besides IWN, within the project were developed: a treebank with a three level syntactic and semantic annotation, a system for integrating NL processors in applications for managing grammatical resources, a dialogue annotated corpus for applications of advanced vocal interfaces, software and tools for advanced vocal interfaces.

A rich linguistic model was designed (Alonge et al. 1998) containing a large set of lexical-semantic relations which are encoded for various subsets of Italian nouns, verbs and adjectives. However the relations systematically encoded are synonymy and hyp(er)onymy. All the synsets³ are also linked to WordNet, the Princeton Wordnet database (Miller et al. 1990).

PAROLE/SIMPLE/CLIPS (henceforth PSC⁴), a four-level lexicon, was elaborated over three different projects: the morphological and syntactic models and lexicons, in the EU LE-PAROLE project; the semantic model and lexicon, in the EU LE-SIMPLE project; the phonological description level as well as the extension of the lexical coverage, in the framework of the Italian national project *Corpora e Lessici dell'Italiano Parlato e Scritto* (CLIPS).

PSC (Ruimy et al., 2002) 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 level⁵, all in full accordance with the international standards set out in the PAROLE-SIMPLE model (Ruimy et al., 1998; Lenci et al., 2000). The theoretical model underlying the semantic representation is based on the EAGLES recommendations and on a revised version of Pustejovsky's Generative Lexicon (Pustejovsky, 1995).

The possibility of using IWN and PSC together, hence taking advantage of the expressive power of both underlying models, is the main goal of the linkage we are planning. A first survey was done (Roventini et al. 2002), which evidenced advantages and problems arising from an actual linkage of these resources. In a further step (Ruimy & Roventini, 2005), an exhaustive comparison of the ontologies allowed to deem a semiautomatic linkage feasible on the whole. According to these studies the major problem is related to the second order entities, which turned out to be often not easily linkable and need a deeper analysis. As shown in (Ruimy &

³Each synset is constituted by various synonyms gathered according to the weak definition of synonymy adopted in WordNet and consequently in IWN, stating that "two expressions are synonymous in a linguistic context C if the substitution of one for the other in C does not alter the truth value" (Miller et al., 1990).

⁴'PSC' is not the acronym of the lexicon and is only used here for the sake of brevity.

⁵Semantic coding was performed in collaboration with THAMUS, Italian Consortium for Multilingual Documentary Engineering.

Roventini, 2005), we may however exploit the partial mappability of the ontological information for automating the linking. IWN could benefit by the syntactic information encoded in PSC, thus gaining a rich syntactic and semantic subcategorisation, by the extensive domain coding, qualia relations, etc. . . . On the other hand, PSC could take advantage by the extensively encoded synonymy and taxonomy relations of IWN. Furthermore, another advantage for PSC could be the possibility of being related to WordNet through the IWN mapping, thus gaining a multilingual dimension. Finally, last but not least, both lexicons would gain in coherence and consistency. The linking process can in fact be considered as a sort of reciprocal evaluation of the two resources, and this is particularly important in this field, where subjectivity is hardly avoidable despite the availability of criteria for coding the lexicon.

In the following paragraphs we describe: i) the main features of these resources; ii) the semantic coding in both lexicons; iii) the analysis carried out on a set of verbal entries; iv) the results and future work.

1 The resources main features

There are a few important differences between these lexicons:

- they are structured in terms of a different type ontology – even though partially mappable: PSC is based on a multidimensional semantic type system organised in a hierarchy, with associated templates of information providing the semantics of the types, while IWN has a set of rather flat top semantic features;
- the basic unit to which all the information is related in PSC is the Semantic Unit (*SemU*), which encodes a single word sense, while in IWN it is the *Synset*⁶;
- PSC is a lexicon strongly structured by means of templates which ensure a basic coherence of coding;
- IWN is larger compared to PSC (67,000 variants⁷ vs. 55,000 senses) and shows a rich, but less systematic and homogeneous representation of information.

In general we find a different philosophy inspiring these lexicons according to the different theoretical models they refer to: WordNet and the Generative Lexicon. In IWN the richness of sense distinctions and the variety of semantic relations holding among the synsets is put in the foreground while PSC's outstanding features are a rich description of the argument structure and selectional preferences of predicative entries, and the connection between syntactic and semantic information.

A characteristic of IWN inherited from the WordNet model is, in particular for verbal entries, the proliferation

⁶This last difference has important consequences in a multilingual environment, e.g. for machine translation. It is in fact not always the case that the variants in the synset are interchangeable translations in any context.

⁷Within a synset we find word senses, or multiwords or also acronyms, of the same PoS, named *variants* of the synset, according to the EWN terminology.

of slightly different senses associated with a lexical item. In PSC, by contrast, more generic and less numerous senses are encoded for a lemma.

On the whole, the two approaches are however not contradictory but just different and in both of them the backbone for lexical representation is provided by an ontology of semantic types and a set of semantic features and relations. We are therefore convinced that a merging of these views is indeed feasible and would be useful and productive for NLP applications.

2 The verb semantics coding

2.1 Verb coding in IWN

2.1.1 The semantic relations

Taking as models both Princeton WordNet and Cruse's approach (Cruse 1986) to meaning representation, a relational view of the lexicon was assumed in EWN and then in IWN according to which all the semantic aspects regarding the lexical level are reflected in the paradigmatic and syntagmatic relations obtaining between any two words in a language.

Therefore, the meaning of a word is described both in terms of other words displaying a similar meaning in a specific context (or synonymous words grouped together within a *synset*) and by referring to the relations that a word has with other words in the lexicon, i.e. to its location within a net. Many lexicalization patterns of 'semantic components' were encoded, whenever possible, without drawing a sharp distinction between what is strictly speaking 'semantic' and what could be described as 'pragmatic meaning'.

This can be seen in particular in the verb coding, where the INVOLVED relation is used to encode data on arguments or adjuncts lexicalized within the meaning of a verb. This relation links a verb and a first order noun whose meaning is 'strongly' connected with the verb itself⁸. Specific subtypes of this relation (AGENT, PATIENT, INSTRUMENT, LOCATION) have also been chosen by taking into consideration results of theoretical research on these notions.

The CAUSE relation is instead used to connect different second order entities to each other. It holds between verbs or links verbs to adjectives or to nouns denoting events or processes. This relation as well has specific subtypes (RESULTS_IN, FOR_PURPOSE_OF, IS_MEANS_FOR).

In Table 1 the rich set of semantic relations encoded for verbs is listed, except for synonymy (which is used to identify synsets).

However, within the IWN verbs net, an uneven distribution of encoded semantic relations was determined by the fact that only the hyp(er)onym and the equivalence_relation to the ILI were mandatory while other relations were encoded for large sets of verbal entries but not systematically.

2.1.2 The second order ontology

The verbs, as entities belonging to the second order, are organized in two different classification schemes, which repre-

⁸The relation *Role* is used for the opposite link, from concrete nouns to verbs (or nouns referring to states, processes or events).

Table 1: IWN semantic relations for verbs

Relation Name	Type of Entities	Examples
ANTONYM	2°Order/2°Order	to love / to hate - to hate / to love
HAS-HYP(ER)ONYM	2°/2°	to see / to perceive - to perceive / to see
CAUSES/IS_CAUSED_BY	2°/2°	to kill / to die – execute / sentence
RESULTS_IN/IS_RESULT_OF	2°/2°	to kill / to die – sick /to fall ill
FOR_PURPOSE_OF/IS_PURPOSE_OF	2°/2°	to search / to find - to win / to compete
IS_MEANS_FOR/HAS_MEANS	2°/2°	heat / distillation - to evaporate / boiling
HAS_SUBEVENT/IS_SUBEVENT_OF	2°/2°	to buy / to pay - to snore / to sleep
INVOLVED/ROLE	2°/1° - 1°/2°	to hammer/hammer – pedestrian/to walk
INVOLVED_AGENT/ROLE_AGENT	2°/1° - 1°/2°	to teach / teacher – runner / to run
INVOLVED_PATIENT/ROLE_PATIENT	2°/1° - 1°/2°	to teach / student – student / to teach
INVOLVED_INSTRUMENT/ROLE_INSTRUMENT	2°/1° - 1°/2°	to paint / paint-brush – gun / to shoot
INVOLVED_LOCATION/ROLE_LOCATION	2°/1° - 1°/2°	to swim / water – school / to teach
INVOLVED_DIRECTION/ROLE_DIRECTION	2°/1° - 1°/2°	to lead / place – arrival / to arrive
INVOLVED_SOURCE_DIRECTION/ ROLE_SOURCE_DIRECTION	2°/1° - 1°/2°	to disembark / ship – outside / to enter
INVOLVED_TARGET_DIRECTION/ROLE_TARGET_DIRECTION	2°/1° - 1°/2°	to exit / outside – inside / to enter
INVOLVED_RESULT/ROLE_RESULT	2°/1° - 1°/2°	to freeze / ice – ice / to ice
IN_MANNER/MANNER_OF	2°/2°	to whisper / in a low voice
XPOS_NEAR_SYNONYM	2°/2°	to arrive / arrival

sent the first division below 2nd Order Entity: *Situation Type* and *Situation Component* (cf. Table 2). The *Situation Type* is connected with the event-structure or *Aktionsart* (lexical aspect) of a situation, while the *Situation Component* lists the most salient semantic components that characterize situations.

Each second order synset is characterized by one well-defined and precise situation type to which many different combinations of *Situation Component* concepts are associated.

2.2 Verbs coding in PSC

In the PSC lexicon, the semantic content of a verb is expressed by its membership of a semantic type (cf. Table 3) which inherently triggers the instantiation of a rich bundle of semantic features and relations. Among these are the 60 relations of the Extended Qualia structure, an enlarged version of the GL representational tool that enables to describe the componential aspect of a word meaning as well as its relationships to other lexical items. The semantic description of verbs also encompasses contextual information, formulated in terms of a semantic predicate and its arguments with their thematic roles and semantic typing. Syntactic and semantic information concerning a verb is linked through the projection of the predicate-argument structure onto its syntactic realization(s).

A basilar element in PSC semantic coding is the *template*, i.e. a schematic structure which allows to constrain a semantic type to a structured cluster of information considered crucial to its definition and eases the lexicographer's task, thus enhancing the consistency and structuring the linguistic information encoded.

As can be seen in the table below the PSC ontology is more structured and detailed compared with the IWN one. This led us to take this ontology as the point of reference

for our semiautomatic linking as we will illustrate in the following paragraphs.

Table 2: IWN second order ontology

2 nd ORDER ENTITY
SITUATION COMPONENT
Cause
Communication
Condition
Existence
Experience
Location
Manner
Mental
Modal
Physical
Material
Physiological
Possession
Purpose
Quantity
Social
Time
Intensity
Property
Attribute
Functional
Relation
SITUATION TYPE
Dynamic
BoundedEvent
UnboundedEvent
Static

Table 3: SIMPLE-CLIPS semantic types for events

Event	
Phenomenon	Weather verb Disease Stimulus
Aspectual	Cause aspectual
State	Exist Relational state Identificational state Constitutive state Stative location Stative possession
Act	Non_relational_act Relational_act Cooperative_activity Purpose_act Move Cause_motion Cause_act Speech_act Cooperative_speech_act Reporting_event Commissive_speech_act Directive_speech_act Expressive_speech_act Declarative_speech_act
Psychological_event	Cognitive_event Judgement Experience_event Cause_experience_event Perception Modal_event
Change	Relational_change Constitutive_change Change_of_state Change_of_value Change_of_possession Transaction Change_of_location Natural_transition Acquire_knowledge
Cause_change	Cause_Relational_change Cause_Constitutive_change Cause_Change_of_state Cause_Change_of_value Cause_Change_location Cause_Natural_transition
Creation	Physical_creation Mental_creation Symbolic_creation Copy_creation Give_knowledge

3 The comparison

The intrinsic complexity of 2nd Order Entities and the differences that emerged when comparing their coding in both resources (cf. Roventini et al., 2003; Ruimy & Roventini 2005) leaves no doubt about the fact that linking verbal entries is a tricky issue. Nevertheless, after an exhaustive mapping of the two ontologies, we are convinced that such a task is worth pursuing. In fact a semiautomatic link of this type of entries and the consequent possibility of a joint consultation is particularly desirable for verbs which, compared to concrete nouns, have much more to gain just for their complexity. For this reason, we decided to enlarge and deepen our analysis comparing sets of verbal entries.

The semiautomatic link we plan to perform should take as reference points both the ‘is-a’ or hyperonymy relation and the ontological concepts. However, given the different philosophy and consequently the different structures underlying the resources we first carried out a punctual, manual test in order to consider and weigh problems and advantages.

We focused our attention on the large set of PSC verbs belonging to the semantic types ‘Experience_Event’ (inchoatives) and ‘Cause_Experience_Event’ (causatives) and, among them, with particular attention we analyzed more than one hundred verbal entries sharing the generic meaning: ‘to feel or to cause a feeling, an emotion a sentiment’ in both resources.

A first problem that this linking will have to deal with, namely the different granularity of sense distinctions, is well illustrated by the handling, in both resources, of the most generic and polysemous verb in the selected field, i.e. *sentire* (to feel, to experience).

We can in fact observe that the various different meanings of this verb are described in IWN in 16

different synsets where *sentire* is combined with other 25 different verbs (out of which 7 verbal multiwords or locutions)⁹; in PSC, by contrast, the same verb has only 9 senses and only 5 synonyms are indicated.

Such a discrepancy between the two lexicons is obviously not always so sharp, but yet a similar situation often occurs. In fact, IWN, following the WordNet model, tends to overdetail the senses and to combine as many synonyms as possible within a synset; PSC, on the other hand, only accounts for fundamental meaning distinctions. However, once this imbalance is ascertained, a good harmonization of the resources should overcome this difference, turning it into an advantage and the fine-grainedness of the PSC ontology should play a fundamental role to achieve in this task.

If we take into consideration the hyperonyms encoded for the various senses of *sentire* in both resources we find a high level of correspondence. As can be seen, comparing tables 4 and 5 below, for 7 out 9 PSC SemUs of *sentire* the corresponding IWN synset shows the same hyperonym and a comparable ontology concept: the PSC semantic type ‘Perception’ matches the IWN ‘Dynamic Experience’ ontological typing, PSC ‘Cooperative_speech_act’ and ‘Ac-

⁹In IWN about 500 multiwords of verbal type were encoded.

quire_knowledge' match the more generic IWN 'Agentive Dynamic' and 'Dynamic Possession' respectively (cf. IWN / SIMPLE_CLIPS ontology mapping table¹⁰). The last PSC meaning, whose hyperonym is *provare* (to experience), has no corresponding hyperonym in table 5. However, its ontological label indicates as possible matches the IWN synsets showing the IWN ontology concepts 'Experience Mental SituationType' which perfectly correspond to PSC 'Experience_event'.

In this case the comparison evidences a fuzzy distinction in IWN between the meanings: 'to experience a feeling, an emotion accompanied by a physical sensation', synset 33902 {*sentire, avere, avvertire, provare*}¹¹ and 'to experience a feeling', synset 33903 {*sentire, provare*}. In PSC the first of these meanings is assigned the semantic type 'Perception' and the second one, 'Experience_event'.

The meanings of *sentire* that share same hyperonym and ontological classification as 'Perception' in PSC and 'Dynamic Experience' in IWN indicate the synonymy of *sentire* with the four perception verbs: *udire* (to hear), *odorare* (to smell), *gustare* (to taste) and *toccare* (to finger). When linked, these entries will acquire the argument structure from PSC and some other synonyms as well as an English translation from IWN.

As regards the meanings not classified as 'Perception' in PSC, we find a good correspondence between the IWN synsets and PSC semUs, e.g.: synset 33896 {*sentire, consultare*} (consult, ask_for_advice) and SemU70608*sentire* share the hyperonym *chiedere* (to ask), while synset 33897 {*sentire, apprendere, venire a sapere, sapere*} (to learn) and SemU70609*sentire* share the hyperonym *acquisire* (to acquire).

The IWN synset 33898 {*sentire, informarsi*} (to ask, to inquire) with hyperonym *interrogare* (to interrogate, to ask questions), which seems at first glance to lack a corresponding SemU, can be equally linked through the inchoative form of SemU79609*informare* that belongs to the semantic type 'Acquire_knowledge' (mappable to 'Agentive Dynamic') and is assigned *acquisire* as target of the hyperonymic relation.

As far as the other IWN synsets are concerned, in a few cases we have very fine distinctions not accounted for in PSC such as, for example, the distinction between {*sentire, udire, intendere*} = 'to hear' with hyperonym *percepire* (to perceive) and {*sentire*} = 'to have the sense of hearing' with hyperonym *potere* (to be able); or between {*sentire, ascoltare*} = 'to listen to', hyperonym *udire* (to hear), and {*sentire, dare ascolto, dare retta, ascoltare*} = 'to obey, to mind', hyperonym *obbedire* (to obey).

In all, we can link 10 out 16 IWN synsets to the corresponding PSC semUs. For every linked verbal entry, the two resources will offer more information: IWN will enrich PSC entries with an English translation and synonyms, also in form of multiwords, and PSC will provide IWN synsets

¹⁰http://www.ilc.cnr.it/clips/Ontology_mapping.doc

¹¹The verbs gathered between brace brackets represent the complete IWN synsets that for space reasons were not inserted in table 5 in their wholeness.

with the argument structure and an undoubtedly more precise ontological classification.

Considering the whole set of analysed senses we can make further observations about: i) the coding of feeling denoting verbs; ii) the results of the comparison between hyperonyms and ontology concepts in both databases.

In the example below, the IWN semantic coding of the figurative senses (both inchoative and causative) of the verb *abbattere* (to lose hope, to despair/'to dishearten, to depress) are shown and then compared to their PSC coding.

inchoative sense ◇ {*abbattersi, abbacchiarsi, accasciarsi, avvilitarsi, demoralizzarsi, deprimersi, scoraggiarsi, sgomentarsi*}

- Has_hyperonym: *diventare* (become)
- xpos_near_synonym: {*abbattimento avvilitamento sconforto abbacchiamento prostrazione depressione*}
- Ontology: BoundedEvent
- eq_near_synonym: (despair, abandon_hope, lose_hope) (ILI link)

causative sense ◇ {*abbattere, accasciare, buttare giù, deprimere, demoralizzare, avvilitare, abbacchiare, sconfortare*}

- Has_hyperonym: *scoraggiare* (discourage)
- Causes: {*abbattimento avvilitamento sconforto abbacchiamento prostrazione depressione*}
- Ontology: Cause
- eq_near_synonym: (dishearten, put_off), (depress, deject, cast_down, get_down, make_gloomy, dismay, dispirit, demoralize) (ILI link)

If we compare the PSC¹² entries for the same meanings we find essentially these main differences:

- in PSC both the causative and the inchoative meanings of *abbattere* bear a relation indicating their sharing the membership in a class of regular polysemy alternation, while in IWN this relation is not encoded.
- in PSC both senses are displayed with the same spelling, while in IWN the inchoative meaning is characterized, where relevant, by the clitic pronoun *-si* (which is peculiar of this type of alternation).

As far as similarities are concerned, in PSC the causative and inchoative meanings are linked to the corresponding feeling-denoting noun *abbattimento* (prostration, demoralization) by means of the semantic relations 'has_as_effect' and 'feeling' which are comparable to 'cause' and 'xpos_near_synonym' respectively.

In IWN as well both the verbal synsets are connected to the corresponding nominal synsets through a 'cause' relation for the causative sense and through an 'xpos_near_synonym' relation for the inchoative sense.

¹²The complete representation of the PSC entries is not given here because of its complexity.

Table 4: Hyperonyms and semantic types in PSC for *sentire*

Hyperonym (isa relation)	Semantic Type	n. of SemUs
<i>percepire</i> (to perceive)	Perception	6
<i>chiedere</i> (to ask)	Cooperative_speech_act	1
<i>acquisire</i> (to acquire)	Acquire_knowledge	1
<i>provare</i> (to experience)	Experience_event	1

Table 5: Hyperonyms and ontology concepts in IWN for *sentire*

Hyperonym	Ontology Concept	n. of Synsets
<i>percepire</i> (to perceive)	DynamicExperience	5
<i>udire</i> (to hear)	DynamicExperience	1
<i>chiedere</i> (to ask)	AgentiveDynamic	1
<i>acquisire</i> (to acquire)	DynamicPossession	1
<i>essere</i> (to be)	ExperienceMental Sit.Type	1
<i>avere</i> (to have)	ExperienceMental Sit.Type	1
<i>potere</i> (to be able)	CauseModalProperty	2
<i>obbedire</i> (to obey)	AgentiveDynamic	1
<i>interrogare</i> (to ask question)	AgentiveDynamic	1
<i>credere</i> (to believe)	MentalProperty	1
<i>accorgersi</i> (to realize)	BoundedEvent	1

The PSC ontological types assigned respectively to the causative and inchoative senses: ‘Cause_experience_event’ and ‘Experience_event’ are comparable to IWN ‘Dynamic Experience Mental Stimulating’ and ‘Experience Mental Situation Type’.

In this particular example, both the IWN inchoative and causative meanings lose some specific feature in their hyperonymic chain and, consequently, the ontological description is more generic – i.e. ‘BoundedEvent’ instead of ‘Experience Mental SituationType’ and ‘Cause’ instead of ‘Dynamic Experience Mental Stimulating’ – but still comparable as observed in (Ruimy & Roventini 2005).

Summing up, as noticed for the verb *sentire*, in this case also the linking of these entries will provide more information and in a more accurate way. In particular, the Polysemy relation, which keeps joint the inchoative and causative word-senses and gives prominence to a typical lexical-semantic rule is provided by PSC, the many synonyms of both senses encoded and the indication of the pronominalization for the inchoative form are supplied by IWN, together with the link to WN 1.5. Furthermore it is worth noting that the relationship between the verb of feeling and the feeling-denoting noun, which is considered optional information in the IWN model, is systematically encoded in PSC.

4 The results and future work

From our analysis it turned out that in IWN causative and inchoative verbs of feeling are not retrievable under a unique hyperonym or a unique ontological concept, but, given the lack of predefined rigid structures for the coding of verbs types, like the templates used in PSC, they can be found under a few different hyperonyms which, in some cases, also point to different ontological concepts, as illustrated in tables 6 and 7 below, where hyperonyms and relative ontology concepts are reported.

As appears in these tables, occasionally we find too much genericity in the hyperonyms coding and, consequently, in the ontological labels, in particular as regards the causative verbs. The survey evidenced that in IWN the hyperonyms of causatives word senses of feeling verbs point, for the most part (40 cases out of 60), to the generic ontology concept ‘Cause’ instead of the more appropriate ‘Dynamic Experience Mental Stimulating’.

On the contrary, as regards the inchoatives senses, the coding appears more precise and the appropriate hyperonym {*provare, sentire*} (to feel, to experience) was encoded for the most part of them. Consequently, the correct ontology label ‘Experience Mental Situation Type’ was found in 30 cases out of 36.

Summing up we found that 44 synsets out of 96 show a perfect correspondence with PSC entries and that all the others show a more generic but not misleading coding, except for the last hyperonym in table 7. This is an encouraging result and, on this basis, we intend to complete the comparison in a semiautomatic way.

Given the different number of verbal entries (5260 SemUs and 9301 synsets) and the greater homogeneity of coding guaranteed by the PSC templates, we will proceed in the comparison extracting the various verbal SemUs from PSC. One semantic type at a time all the verbal SemUs of PSC will be matched to the corresponding synsets in IWN. Once completed the automatic extraction of the matched couples, the candidate joint entries will be checked for mistakes. By means of this procedure we expect to be able to link nearly all the verbal PSC SemUs with a corresponding IWN synset, and to circumscribe in this way a set of verbs showing joint together the most valuable features of both resources.

Table 6: IWN Hyperonyms and Ontology concepts of causative verbs of feeling

Hyperonyms	n. of Synsets	Ontology Concepts
<i>Rendere</i> 2 (make 1 get 1)	16	Cause
<i>Fare</i> 14 (cause 1 induce 1 stimulate 1..)	12	Cause
<i>Dare</i> 2 (give3 cause to have 1)	8	Cause
<i>Accorare</i> 1 (grieve 1 aggrieve 1 sadden_greatly 1)	1	Cause
<i>Agitare</i> 3 (agitate 1 commove 1 excite 4 ..)	1	Cause
<i>Fare</i> 11 (cause 1 do 3 make 9 give_raise_to 1)	1	Cause
<i>Scoraggiare</i> 1 (discourage 1)	1	Cause
<i>Suscitare</i> 1 (arouse 1 elicit 1 enkindle 1 fire 1..)	5	Dyn.Exp.MentalStimulating
<i>Incutere</i> 1 (instill 1)	4	Dyn.Exp.MentalStimulating
<i>Fare</i> 11 (cause 1 do 3 make 9 give_raise_to 1)	3	Dyn.Exp.MentalStimulating
<i>Causare</i> 2 (arouse 1 elicit 1 enkindle 1 fire 1..)	1	Dyn.Exp.MentalStimulating
<i>Impaurire</i> 1 (frighten 1 fright 1 scare 1..)	1	Dyn.Exp.MentalStimulating
<i>Fare</i> 11 (make 1 create 1)	5	CauseExistence
<i>Rendere</i> 2 (make 1 get 1)	1	CauseExistence

Table 7: IWN Hyperonyms and Ontology concepts of inchoative verbs of feeling

Hyperonyms	n. of Synsets	Ontology Concepts
<i>Sentire</i> 10 (feel 1 experience 1)	23	Exp.Mental Sit.Type
<i>Avere</i> 2 (have 1 have_got 1 hold 1)	2	Exp.Mental Sit.Type
<i>Essere</i> 1 (be 4 have_the_quality_of_being 1)	2	Exp.Mental Sit.Type
<i>Provare</i> 6 (experience 1 undergo 1 get 1 have 6..)	2	Exp.Mental Sit.Type
<i>Desiderare</i> 1 (desire 1 want 1 wish 1)	1	Exp.MentalSit.Type
<i>Diventare</i> 1 (become 1 go 5 get 1 come 1 ..)	2	BoundedEvent
<i>Incollerire</i> 1 (anger 1 get_angry 1, see_red 1)	1	BoundedEvent
<i>Subire</i> 1 (experience 1 undergo 1 get 1 have 6..)	1	DynamicExperience
<i>Prendere</i> 10 (experience 1 undergo 1 get 1 ..)	1	DynamicExperience
<i>Essere</i> 1 (be 4 have_the_quality_of_being 1)	1	Static

5 Conclusion

In this paper we described a detailed preliminary analysis aimed at investigating the possibilities of semi-automatically linking the two largest and richest Italian lexical resources, IWN and PSC, as far as second order entities are concerned. The methodology adopted, which is grounded on the mapping of both hyperonymy relations and ontological classification, turned out to be sufficiently reliable and the preliminary results promising enough to encourage us to carry on the linking process.

References

- Alonge, A., Calzolari, N., Vossen, P., Bloksma, L., Castellon, I., Marti, T., Peters, W. (1998), The Linguistic Design of the EuroWordNet Database, in “Computers and the Humanities”, N. Ide, D. Greenstein (eds.), Special Issue on EuroWordNet, Guest Editor: Piek Vossen, XXXII, 2–3, 91–115.
- Lenci, A., Bel, N., Busa, F., Calzolari, N., Gola, E., Monachini, M., Ogonowsky, A., Peters, I., Petres, W., Ruimy, N., Villegas, M., Zampolli, A., (2000) SIMPLE: A General Framework for the Development of Multilingual Lexicons, in “International Journal of Lexicography”, XIII 4: 249–263.
- Miller, G., Beckwith, R., Fellbaum C., Gross D., Miller K.J. (1990) *Introduction to WordNet: An On-line Lexical Database*, “International Journal of Lexicography”, III, 4, 235–244.
- 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*, in “Linguistica Computazionale”, A. Zampolli, N. Calzolari, L. Cignoni, (eds.), vol. XVIII-XIX, Pisa-Roma, IEPI. Tomo II, pp. 745–791.
- Roventini, A., Ulivieri, M., Calzolari, N. (2002) *Integrating two semantic lexicons, SIMPLE and ItalWordNet: what can we gain?*, in “LREC Proceedings of the Third International Conference of Language Resources and Evaluation”, Vol. V, pp. 1473–1477.
- Ruimy, N., Battista, M., Corazzari, O., Gola, E., Spanu, A. (1998) *Italian Lexicon Documentation, LE-PAROLE, Final Review Documentation, LE-4017 Parole Project*.
- 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 “Linguistica Computazionale”, A. Zampolli, N. Calzolari, L. Cignoni, (eds.), vol. XVIII–XIX, Pisa-Roma, IEPI. Tomo II, 821–864.
- Ruimy, N., Roventini, A. (2005) *Towards the Linking of two Electronic Lexical Databases of Italian*, in Zygmunt Veutulani (ed.), L&T’05 – 2nd Language Technologies

as a Challenge for Computer Science and Linguistics,
April 21–23, Poznan, Poland. Wydawnictwo Poznanskie
Sp. z o.o., 230–234.
Vossen, P. (ed.) (1999): EuroWordNet General Document,
<http://www.hum.uva.nl/~EWN>.