Mapping Events and Abstract Entities from PAROLE-SIMPLE-CLIPS to ItalWordNet

Adriana Roventini, Nilda Ruimy
Istituto di Linguistica Computazionale, CNR
Via Moruzzi 1 – Pisa - Italy
{adriana.roventini, nilda.ruimy}@ilc.cnr.it

Abstract

In the few last years, due to the increasing importance of the web, both computational tools and resources need to be more and more visible and easily accessible to a vast community of scholars, students and researchers. Furthermore, high quality lexical resources are crucially required for a wide range of HLT-NLP applications, among which word sense disambiguation. Vast and consistent electronic lexical resources do exist which can be further enhanced and enriched through their linking and integration. An ILC project dealing with the link of two large lexical semantic resources for the Italian language, namely ItalWordNet and PAROLE-SIMPLE-CLIPS, fits this trend. Concrete entities were already linked and this paper addresses the semi-automatic mapping of events and abstract entities. The lexical models of the two resources, the mapping strategy and the tool that was implemented to this aim are briefly outlined. Special focus is put on the results of the linking process: figures are reported and examples are given which illustrate both the linking and harmonization of the resources but also cases of discrepancies, mainly due to the different underlying semantic models.

1. Introduction

Nowadays, when building lexical resources much attention is paid to their use in HLT-NLP applications and their interoperability. In fact, given the relevant development of cross-lingual studies and applications it is of utmost importance that linguistic information be easily accessible, comparable and manageable by common software and flexible computational tools. High quality lexical resources are crucially required for a wide range of applications, among which word sense disambiguation. Vast and consistent electronic lexicons do exist which can be further enhanced and enriched through their linking and integration. Linking and integrating lexical resources is in fact a trend that is getting more and more attention. A combination of WordNet, FrameNet and VerbNet was, for example, recently proposed in order to enhance the semantic parsing as far as English is concerned (Shi and Mihalcea, 2005).

In this perspective, the link between PAROLE-SIMPLE-CLIPS (Ruimy *et al.*, 2003) and ItalWordNet (Roventini *et al.*, 2003), the two largest and extensively encoded computational lexicons of Italian, appeared to be a choice in line with a new generation of language resources but also fruitful for the reciprocal benefits expected. IWN will in fact benefit from an exhaustive syntactic information and the description of semantic predicates provided by PAROLE-SIMPLE-CLIPS; the latter will, in turn, be enhanced with the rigorous taxonomy relations, the rich synonymy encoding and the link to Princeton WordNet provided by IWN. The two resources, although structured according to different lexical models, present many compatible aspects which were considered a good starting point to carry out their link.

Yet, the linking process of these large and complex lexical resources was not a trivial task and it has been devised and carried out over various phases. First, a feasibility study about linking — at least partially automatically — the two lexicons evidenced a few problematic issues but also the many advantages related to this operation (Roventini et al., 2002). Then, a comparison of the lexical models underlying the resources and the mapping of their ontological framework was performed[1], semantic relations were compared and a linking methodology was devised (Ruimy and Roventini, 2005). Finally a software tool was implemented to map the lexical units of both lexicons and the mapping of all concrete entities was achieved (Ruimy et al., 2008).

At the same time, mapping experiments were manually carried out on sets of verbal entries, which were always considered the most difficult items to deal with. In particular, speech act and feeling verbs were analysed in both resources. Encouraging results led us to tackle the semi-automatic mapping of verbs (Roventini and Ruimy, 2006; Roventini, 2006). In the following sections, the lexical models of the two resources, the mapping strategy and the software tool are briefly outlined; the mapping of event-denoting verbs and nouns and of abstract entities is described and the first results are provided. Finally, some ongoing applications of the mapping are mentioned.

2. The Lexical Models

ItalWordNet (IWN) is a lexical semantic database based on EuroWordNet (EWN) lexical model (Vossen, 1999) which, in its turn, is inspired from Princeton WordNet (Fellbaum, 1998). IWN is organized around the notion of *synset*, i.e. a set of synonymous word-senses or *synset variants*. All IWN synsets are linked to WordNet through an interlingual index (ILI) which makes the resource usable in multilingual applications.

PAROLE-SIMPLE-CLIPS (PSC) is a four-layered lexicon providing phonological, morphological, syntactic and semantic information. It is based on EAGLES recommendations, on the results of EuroWordNet and ACQUILEX projects and on a revised version of Pustejovsky's Generative Lexicon theory (Pustejovsky, 1995). At the semantic level (referred to as 'SIMPLE-CLIPS', from now on), the basic unit is the word sense, represented by a 'semantic unit' (*SemU* or *USem* [2]).

In both lexicons, the main structure for lexical representation is provided by an ontology of semantic

[1] http://www.ilc.cnr.it/clips/Ontology_mapping.doc
[2] English and Italian acronyms.

types (SIMPLE-CLIPS) or top concepts (IWN).

IWN Top Ontology (TO) consists of a hierarchical structure of 65 language-independent *Top Concepts* (TCs) clustered in three categories distinguishing 1st Order Entities, 2nd Order Entities and 3rd Order Entities. Their subclasses, hierarchically ordered, are also structured in terms of (disjunctive and non-disjunctive) opposition relations. Each synset is ontologically classified on the basis of its hyperonym and, in most cases, cross-classified in terms of multiple, non disjoint TCs, e.g.: *cardiologia* (cardiology) [Agentive Purpose Social Undboundedevent].

SIMPLE Ontology (SO)[3], which consists of 157 language-independent semantic types, is a multidimensional ontology based on hierarchical and non-hierarchical conceptual relations. In the type system, multidimensionality is captured by *qualia roles* that define the distinctive properties and differentiate the internal semantic constituency of semantic types. SO distinguishes therefore between *simple* (one-dimensional) and *unified* (multi-dimensional) semantic types, the latter implementing the principle of orthogonal inheritance (Pustejovsky and Boguraev, 1993).

3. Semantic Representation

In IWN, the meaning of a word is described both in terms of other lexical items displaying a similar meaning in a specific context and by referring to its relations with other words in the lexicon, i.e. to its location within a net. Many lexicalization patterns of 'semantic components' were also encoded, whenever possible. For example, for verbs, the *involved* relation is used to encode data about arguments or adjuncts lexicalized within the meaning of a verb. This relation links a verb and a 1st Order noun whose meaning is connected with the verb itself[4]. Specific subtypes of this relation (*agent, patient, instrument, location*) make it particularly useful.

Events, which belong to the Second Order (Lyons, 1977), are organized in terms of two different classification schemes which represent the first division below 2nd Order Entities: *Situation Type* and *Situation Component*. The *Situation Type* is connected with the event-structure or Aktionsart (lexical aspect) of a situation. Two different aspects are distinguished: *Static* and *Dynamic* (which, in turn, has as subtypes *BoundedEvent* and *UnboundedEvent*). The *Situation Component* lists 22 salient semantic components that characterize situations.

Abstract entities, existing independently of time and space, belong to 3rd Order Entities.

In the SIMPLE-CLIPS database, the semantic content of a word sense, be it a concrete/abstract entity or an event, is expressed by the semantic type it belongs to and by a rich bundle of semantic features and relations entering in the definition of this type. The use of *templates*, i.e. schematic structures which allow to constrain a semantic type to a structured cluster of information considered crucial to its definition, enables a consistent structuring of information. Among the expressive means for the representation of meaning are the 60 relations of the *Extended Qualia Structure* that enable to describe the componential aspect of a word meaning as well as its relationships to other lexical items.

For all predicative units, the semantic description also includes contextual information. Event-denoting entries are therefore connected to a lexical predicate which is described in terms of arity, thematic role and semantic constraints of each semantic argument. The predicate-argument structure is projected onto its syntactic realization(s), thus ensuring the link between syntactic and semantic information.

The event structure is expressed, in the SIMPLE-CLIPS database, by means of the three-valued feature *Eventtype = state*, *process*, *transition*, values which correspond in IWN to the *Situation Type* values *Static*, (*Dynamic*) *Unboundedevent* and (*Dynamic*) *Boundedevent,* respectively.

As to the IWN *Situation Component* parameter, it is impossible to establish a precise correspondence among the many and various IWN combinatorial subclasses and SIMPLE types. Each semantic component characterizing a situation generally corresponds to one or more SIMPLE types, depending on the value of the *Situation type* and on the other *Situation components* it combines with.

Concerning abstract entities, they are structured in terms of a eight-type sub-hierarchy of the SIMPLE semantic type ABSTRACT_ENTITY.

4. Mapping Process

With its 157 semantic types, SIMPLE ontology allows a more fine-grained structuring of the lexicon than the 65 top concepts of the IWN ontology, which reflect only fundamental distinctions. It has therefore been taken as input for the mapping process which is SIMPLE-CLIPS → IWN oriented.

The lexical instances of a SIMPLE semantic type, along with their PoS and 'isa' relation are taken as starting point. The IWN resource is then explored in search of linking candidates with same PoS and whose ontological classification matches the correspondences established between the classes of both ontologies[5].

In the following table examples are given of some corresponding ontological classifications.

[3] http://www.ilc.cnr.it/clips/Ontology.htm

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

[5] see note 1.

SIMPLE semantic type	IWN TCs combinations
ACT	Agentive Dynamic UnboundeEvent
CAUSE_CHANGE	Cause BoundedEvent
METALANGUAGE	3rd Order Mental Language Representation

Table 1. Corresponding ontological tags

During the linking process, hyperonyms of matching candidates are also taken into account and play a particularly determinant role in the resolution of cases whereby matching fails, due to a conflict of ontological classification.

Concretely, the Access software tool implemented to map the lexical units of both lexicons works in a semi-automatic way using the ontological classifications, the 'isa' relations and some semantic features of the two resources. The mapping process foresees the following steps:

- selection of a PSC semantic type and definition of the loading criteria, i.e. either all its *SemUs* or only those bearing a given information;
- selection of one or more mapping constraints on the basis of the correspondences established between the conceptual classes of both ontologies, in order to narrow the automatic mapping;
- human validation of the automatic mapping and storage of the results;
- if necessary, relaxation/tuning or addition of mapping constraints and new processing of the input data.

Human validation of the automatic mapping consists in the manual selection of the semantically relevant word sense pair(s), when more than one possible match between a *SemU* and a *synset variant* is automatically output (this is referred to as *multiple mapping,* see table 2 below). Human validation then occurs after checking relevant information sources such as hyperonyms, *SemU* and *synset* glosses and the IWN-ILI link.

SIMPLE-CLIPS		IWN			
Usem	Semantic Type	Synset Variant	Synset number	Sense	TCs
USem60693affliggere	CAUSE_EXPERIENCE_EVENT	affliggere	34763	1	Cause
USem60693affliggere	CAUSE_EXPERIENCE_EVENT	affliggere	36948	2	Cause
USem60693affliggere	CAUSE_EXPERIENCE_EVENT	affliggere	35011	3	Cause

Table 2. An example of multiple mapping

Besides the matched pairs of word senses, i.e. *SemUs* and *synset variants* showing identical string, PoS and comparable ontology codes, each mapping run returns:

- a list of unmatched words containing the IWN word senses whose *synset* ontological classification is incomplete or different w.r.t. the constraints imposed to the mapping run;
- a list of SIMPLE-CLIPS lexical units or lemmas missing in IWN.

The first list is crucial to identify further mapping constraints, as it provides a statement of the discrepancies regarding the ontological classification in the two lexicons. The second one informs on the lexical intersection between the resources.

5. Mapping Events & Abstract Entities

The different philosophy inspiring these lexicons and, above all, the different granularity in both word sense distinction and ontological classification appeared even more evident when addressing the linking of events and abstract entities. In fact, the characteristic proliferation of slightly different senses of IWN verbal entries made the human validation of the automatic mapping more difficult and problematic. As shown in table 2 above, in many cases, a unique *SemU* could be linked to different *synsets* (here, three slightly different senses of the verb *affliggere*, with identical ontological classification) and the lexicographer had therefore to validate one or more matches from the output of multiple mappings[6].

[6] Note that the lists of multiple mapping constitute a useful repository available for further analyses and (re)considerations.

At the same time, the highly structured SIMPLE ontology imposed in some cases an even too detailed grid of comparison. For typing speech act events, for example, the SIMPLE ontology avails of a hierarchy of six semantic types subsumed by the top type SPEECH_ACT (table 3).

SIMPLE Semantic Types	IWN Top Concepts
SPEECH_ACT	Agentive Communication Dynamic
COOPERATIVE_SPEECH_ACT	Agentive Comm. Unbounded Event
REPORTING_EVENT	Agentive Comm. Dynamic
COMMISSIVE_SPEECH_ACT	Agentive Comm. Dynamic
DIRECTIVE_SPEECH_ACT	Agentive BoundedEvent Comm. Purpose
EXPRESSIVE_SPEECH_ACT	Agentive Comm. Dynamic
DECLARATIVE_SPEECH_ACT	Agentive Comm. Purpose

Table 3 Speech Act classification in both ontologies

To link all speech act *SemUs* to the corresponding IWN synsets, seven mapping runs were therefore carried out, every time using the constraint: [Agentive Communication] as a common denominator able to capture all speech act *synsets*.

To give another example of the same problem originating from IWN side, in order to link the *SemUs* belonging to the semantic type PURPOSE_ACT to the corresponding IWN *synsets*, five different constraints were defined, every time relaxing the search field, from the precise foreseen correspondence PURPOSE_ACT → [Agentive

Purpose Social UnboundedEvent] to the most generic PURPOSE_ACT → [Cause]. It is interesting to notice that, in the end, about fifty different TCs combinations were found against the PURPOSE_ACT semantic type.

Besides these two main kinds of discrepancy, other problems we dealt with are an incomplete or different ontological classification of IWN *synsets* and a few inconsistencies due to a different meaning interpretation made by lexicographers.

By incomplete ontological information we intend those cases where, in IWN, the expected combination of TCs is lacking one of the two classifying parameters, either *Situation type* or *Situation component* for 2nd Order entities, and, for abstract entities, either the indication of membership to 3rd Order Entity or one or more *Situation component* tags.

For example, when mapping the semantic units belonging to the SIMPLE semantic type DISEASE we found that a limited number of word senses matched on the basis of the established correspondence: DISEASE → [Dynamic Phenomenal Physical], used as a constraint in the first mapping run. The analysis of the unmatched word senses evidenced both incomplete and different ontological classification. In Table 4 below, four sets of word senses are shown which exemplify the mismatch cases listed above. In the first two groups, slightly different TCs combinations descend from two different hyperonyms: physical illness [Dynamic Phenomenal Physical] and mental illness [Dynamic Experience Mental]. Instead, the classification of some word senses as [Property] depends on a different meaning interpretation made by the lexicographer. In this case, for a considerable number of synsets, a shift occurred from "process" to "condition" along the hyperonymical chain, which also determined a change from [Event] to [Property] in the ontological classification.

Furthermore, all synsets show an incomplete ontological classification: either S*ituation component* or S*ituation type* coding tags were provided.

In any case, thanks to the more reliable SIMPLE ontological classification and a careful analysis of the unmatched output list, cases of seeming incompatibility and 'reasonable' incompleteness were detected and linked.

SIMPLE-CLIPS			IWN			
Usem	**Semantic Type**	**Isa Relation**	**Synset Variant**	**PoS**	**S.**	**TCs**
USem78249ictus	DISEASE	UsemD63847lesione	ictus	N	1	Dynamic Phenomenal Physical
USem68392infarto	DISEASE	UsemD63847lesione	infarto	N	1	Dynamic Phenomenal Physical
USem3831morbillo	DISEASE	Usem3868malattia	morbillo	N	1	Dynamic Phenomenal Physical
USemD5896cleptomania	DISEASE	UsemD7206mania	cleptomania	N	1	Dynamic Phenomenal Mental
USem69149depressione	DISEASE	Usem67631malattia	depressione	N	5	Dynamic Phenomenal Mental
USem74107amnesia	DISEASE	UsemD5439affezione	amnesia	N	1	Dynamic Phenomenal Mental
USem3800acne	DISEASE	Usem3868malattia	acne	N	1	Property
USem3823influenza	DISEASE	Usem3868malattia	influenza	N	2	Property
USem5275dermatite	DISEASE	UsemD5655infiammazione	dermatite	N	1	UnboundedEvent
USemTH 08351eritema	Disease	Usem67631malattia	eritema	N	1c	Dynamic

Table 4. Situation found when mapping DISEASE word senses

As regards the mismatches originated by a different ontological classification, a curious case regards the set of imaginary creatures such as spirits, ghosts, fabulous animals etc., which in PSC are considered abstract entities and assigned to the semantic type REPRESENTATION while in IWN they belong to 1st Order Entities and are classified under the TC [Creature]. All these word senses were linked, since we considered that this discrepancy reflects two different but defensible points of view on a set of lexical items.

Many examples then can be cited of unmatched word senses due to cases of unbalanced polysemy assessment. For example, nouns indicating in their basic meaning a food, fruit, natural substance, flower, animal, precious stone etc. show a colour sense shifting encoded in SIMPLE-CLIPS under the semantic type COLOUR, but have no correspondent in IWN. In fact, only 24 out of the 97 COLOUR-typed *SemUs* were linked to corresponding synsets (see Table 10). This is another typical case of enhancement in case of merging of the resources.

Another difference evidenced by the mapping of verbal entries concerns the verbs showing a causative / inchoative alternation. In SIMPLE-CLIPS, the *SemUs* for both senses display the same spelling[7], while in IWN the inchoative meaning is mostly characterized by the clitic pronoun *–si* (which is peculiar of this type of alternation). Therefore, while causative word senses are linkable, inchoative ones could not be linked in most of the cases. For example, the semantic types CAUSE_CHANGE_OF_STATE and CHANGE_OF_STATE show a linking percentage of 96,17% and 18,65% respectively (see Table 10). This difference could be harmonized in the merging phase.

Summing up, this linking operation, while evidencing a few conflicting and critical points between the resources, allowed nevertheless to enhance their consistency as it implied a reciprocal assessment of both their coverage and accuracy, which is relevant to hand-built lexical

[7] The pronominal spelling, when existing, of the inchoative form is provided somewhere else in the entry.

resources. Cleaning and harmonizing the two lexical resources as regards word sense distinction, ontological typing and polysemy treatment constitutes moreover a step forward towards their interoperability and eases their eventual merging.

6. Mapping Result

Not surprisingly, this last working phase which addressed the mapping of event-denoting nouns and verbs as well as the one of abstract entities was the most complex one. Nevertheless, on the whole, the mapping was completed with good results.

The mapping of all lexical units being just over, no in depth analysis and discussion of the results has been done so far. In the following, we will therefore limit ourselves to providing some figures. First, numerical data will be provided about each mapping phase (concrete entities in table 5, abstract entities and property denoting nouns in table 6 and 7 respectively, events in table 8, all entities in table 9); then, in table 10, the results of this last mapping phase will be shown, with all input semantic types ordered according to the percentage of linking of their semantic units.

For the moment, we will only point out that 69,59% of SIMPLE-CLIPS *SemUs* denoting events and abstract entities were linked to IWN *synsets*. Compared to concrete entities, we observe a quite expectable lower linking percentage and a higher number of multiple mappings, which is due to the greater intrinsic complexity of the categories under study. The mapping of a single *SemU* to *synset variants* from different synsets (see section 4.) explains the percentages over 100% reported in table 10. Another noticeable difference regards the lexical overlapping which is highest in this part of the lexicon.

Selected USems	27,768
Linkable senses	15,193
Linked senses	10,988
Multiple mappings	1,125
Overlapping coverage	54,71%
Linking percentage	**72,32%**

Table 5 Concrete entities mapping results

Selected USems	4,372
Linkable senses	2,723
Linked senses	1,901
Overlapping coverage	62,28%
Linking percentage	**69,58%**

Table 6 Abstract entities mapping results

Selected USems	1,892
Linkable senses	1,003
Linked senses	894
Overlapping coverage	53,01%
Linking percentage	**89,13%**

Table 7 Property denoting nouns mapping results

89,13%Selected USems	10,584
Linkable senses	8,118
Linked senses	6,322
Overlapping coverage	76,70%
Linking percentage	**77,87%**

Table 8 Event-denoting verbs and nouns mapping results

Selected USems	43,346
Linkable senses	28,190
Linked senses	20,468
Multiple mappings	3,124
Overlapping coverage	65,03%
Linking percentage	**72,60%**

Table 9 Global results

SIMPLE-CLIPS Semantic type	***USem***	**Linkable *USem***	**Linked *USem***	**Linking Percentage**
Physical_property	104	102	126	123,52%
Give_knowledge	65	64	71	110,95%
Cooperative_speech_act	60	58	64	110,34%
Physical_creation	42	41	44	107,31%
Reporting_event	62	62	66	106,45%
Copy_creation	31	28	29	103,57%
Social_property	7	5	5	100%
Speech_act	142	134	134	100%
Commissive_speech_act	10	10	10	100%
Declarative_speech_act	15	14	14	100%
Judgement	27	27	27	100%
Cause_constitutive_change	119	114	114	100%
Directive_speech_act	65	64	63	98,43%
Property	254	128	125	97,655
Mental_creation	69	68	66	97,05%
Expressive_speech_act	116	113	109	96,46%
Cause_change_of_state	714	679	653	96,17%
Cause_aspectual	45	45	35	94,59%
Quality	1322	603	561	93,03%
Act	238	223	205	91,92%
Disease	1413	580	523	90,17%
Unit_of_measurement	119	89	80	89,88%
Symbolic_creation	185	178	159	89,32%
Physical_power	9	9	8	88,88%
Cause_change_of_value	77	74	65	87,83%
Information	371	285	249	87,36%
Cause_experience_event	213	208	181	87,01%
Creation	28	27	23	85,18%
Exist	15	13	11	84,61%

Weather_verb	19	19	16	84,21%
Stative_possession	12	12	10	83,33%
Cause_act	78	76	63	82,89%
Stimulus	56	52	43	82,69%
Cause_change	163	155	128	82,58%
Phenomenon	701	394	323	81,97%
Cause_relation_change	62	61	50	81,96%
Cause_natural_transition	30	29	22	81,48%
Relational_act	839	798	625	78,32%
Moral_standard	48	48	37	77,83%
Movement_of_thought	684	374	272	72,72%
Cognitive_event	216	211	161	76,30%
Shape	77	59	45	76,27%
Metalanguage	258	166	126	75,90%
Cooperative_activity	88	85	64	75,29%
Number	126	40	30	75%
Domain	1210	665	497	74,73%
Acquire_knowledge	45	45	32	71,11%
Natural_transition	24	24	17	70,83%
Cognitive_fact	80	64	45	70,31%
Change_of_location	122	118	80	67,79%
Time	323	231	154	66,66%
Transaction	76	76	49	64,47%
Purpose_act	1175	1125	660	58,66%
Abstract_entity	119	115	67	58,26%
Institution	143	129	74	57,36%
Identificational_state	41	40	23	57%
Abstract_location	37	22	12	54,54%
Experience_event	394	385	210	54,54%
Non_relational_act	233	224	121	54,01%
Change_of_value	50	50	27	54%
State	379	367	194	52,86%
Move	220	203	106	52,21%
Representation	265	195	100	50,76%
Change	94	91	46	50,54%
Event	331	314	157	50%
Stative_location	70	69	34	49,27%
Relational_state	52	52	25	48,07%
Language	223	156	67	42,94%
Aspectual	39	37	15	40,54%
Cause_motion	52	52	21	40,38%
Convention	241	192	74	38,54%
Psychological_event	250	239	88	36,82%
Change_of_possession	77	74	25	33,78%
Perception	199	197	71	30,04%
Constitutive_state	30	30	9	30%
Modal_event	92	92	24	26,08%
Color	119	97	24	24,74%
Sign	125	92	18	19,56%
Change_of_state	354	343	64	18,65%
Relational_change	27	27	0	
Constitutive_change	34	33	0	

Table 10. Mapping of the event and abstract entities sorted by linking percentage per semantic type

7. Concluding remarks

This paper described the last phase of the linking process of the two largest, general purpose, electronic lexical resources of Italian language: PAROLE-SIMPLE-CLIPS and ItalWordNet.

The differences regarding the nature of linking units, the granularity of sense distinction and the ontological classification are complex issues which were addressed during the entire linking process and, particularly, when dealing with events and abstract entities. Unpredictable and non-systematic ontological typing, due to incomplete or inconsistent encoding, came to foreground and made the validation process sometimes quite difficult. Nevertheless, the good results obtained proved that the linking initiative was a worthwhile effort.

The linking process made it possible to enrich each resource with complementary information types that are peculiar to the other theoretical model. SIMPLE-CLIPS will benefit by the link to WordNet, the richness of sense distinction and the consistency of hierarchical links existing in IWN; the latter will profit from the rich description of argument structure, the connection between syntactic and semantic information and the well structured and reliable SIMPLE ontology.

It is desirable that, in the near future, these complementary characteristics be structured in a common representational framework where all these features are visible and available. With this linking, we set, in fact, the basis for building a new unified and richer lexical-semantic resource (Calzolari, 2007) where the many and various points of strength will be put in the foreground.

The mapping we performed has immediately found some application fields: it is, in fact, being fruitfully used in a number of new projects and researches.

In the framework of the international project NEDO[8], starting from a basic verb list (mainly EWN/IWN *base concepts*), a core lexicon was built and then exported to *LMF* format (Takenobu et al., 2008). Through the mapping, the encoded entries for the IWN base concepts were linked to SIMPLE-CLIPS *SemUs* and automatically gained additional syntactic and semantic information (syntactic behavior, semantic type, semantic relations among senses and predicative representation) extracted from PAROLE-SIMPLE-CLIPS entries.

The mapping between SIMPLE-CLIPS and IWN is also exploited in a project[9] aimed at the alignment of an Italian corpus with the TimeBank corpus (Pustejovsky et al., 2003) In this project the mapping was crucial to provide a correct event classification in Italian according to a previously established correspondence between

[8]Japanese NEDO - International Joint Research Program of the **N**ew **E**nergy and Industrial Technology **D**evelopment **O**rganization. Developing International Standards of Language Resources for Semantic Web Applications.

[9]"Modello per analisi e estrazione di eventi e espressioni temporali in testi italiani di ambito generale sfruttando le risorse linguistiche ItalWordNet e PAROLE/SIMPLE/CLIPS".

SIMPLE Semantic types and event categories in TimeML, (Caselli et al. 2008).

Furthermore, the mapping is being used in a study, performed in the framework of a PhD thesis, which deals with the identification and classification of events and temporal expressions in texts and with the computation of the temporal relations holding among these entities.

8. References

Calzolari N. (2007). Towards a new generation of Language Resources in the Semantic Web vision. In Khurshid Ahmad, Christopher Brewster, Mark Stevenson (Eds.) *Words and Intelligence* II: *Essays in honour of Yorick Wilks*. Dordrecht, Springer, pp. 63-105.

Caselli T., Ide N., Bartolini R. (2008). A Bilingual Corpus of Inter-linked Events. To appear in *LREC08 Proceedings*, Marrakech, Morocco.

Fellbaum C. (Ed.) (1998). *Wordnet: An Electronic Lexical Database*. MIT Press.

Lyons, J. (1977). *Semantics*. London, Cambridge University Press.

Pustejovsky J., Boguraev B. (1993). Lexical knowledge representation and natural language processing. *Artificial Intelligence*, Volume 63, Special volume on natural language processing, Issue 1-2, pp. 193-223.

Pustejovsky J. (1995). *The generative lexicon*. MIT Press.

Pustejovsky, J., P. Hanks, R. Saurí, A. See, R. Gaizauskas, A. Setzer, D. Radev, B. Sundheim, D. Day, L. Ferro and M. Lazo. (2003). The TIMEBANK Corpus. In *Proceedings of Corpus Linguistics*, pp. 647-656.

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.

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.) *Linguistica Computazionale*, vol. XVIII-XIX, Pisa-Roma, IEPI. Tomo II, pp. 745-791.

Roventini A., Ruimy N. (2006). Linking and harmonizing different lexical resources: a comparison of verbal entries in ItalWordNet and PAROLE-SIMPLE-CLIPS. In *GWC-06 Proceedings,* January 22-24-, South Jeju Island, Korea.

Roventini, A. (2006). Linking verbal entries of different lexical resources. In *LREC Proceedings*, CD-ROM, 1710-1715.

Roventini A., Ruimy N., Marinelli R., Ulivieri M., Mammini M. (2007). Mapping Concrete Entities from PAROLE-SIMPLE-CLIPS to ItalWordNet: Methodology and Results. In *Proceedings of the 45th Annual Meeting of the Association for Computational Linguistics*, Companion Volume, Prague, Czech Republic, The Association for Computational Linguistics, pp.. 161-164.

Ruimy N. et al. (2003). A computational semantic lexicon of Italian: SIMPLE. In A. Zampolli, N. Calzolari, L. Cignoni, (Eds.) *Linguistica Computazionale*, 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.

Ruimy N. (2006). A Computational Multi-Layered Italian Lexicon for HLT Applications. In *Proceedings XII EURALEX International Congress, Atti del Congresso Internazionale di Lessicografia*, Torino, 6-9 settembre, volume I, pp. 221-227.

Ruimy, N., Roventini, A., Marinelli, R., Ulivieri, M. (2008). Linking and Integrating two Electronic Lexicons. In N. Ide, A.C. Fang (Eds.), *The First International Conference on Global Interoperability for Language Resources*. City University of Hong Kong.

Shi L. and Mihalcea R. (2005). Putting Pieces Together: Combining FrameNet, VerbNet and WordNet for Robust Semantic Parsing. In *Proceedings of the Sixth International Conference on Intelligent Text Processing and Computational Linguistics*, Mexico, pp..

Takenobu T, Sornlertlamvanish V., Charoenporn T., Calzolari N., Monachini M., Soria C., Huang C.R., Hsieh S.H., Kiyoaki, YingJu X (2008). An infrastructure to enhance language technologis for Asian language. To apper in *LREC08 Proceedings*, Marrakech, Morocco.

Vossen P. (Ed.) (1998). *EuroWordNet: A multilingual database with lexical semantic networks*. Kluwer Academic.