An Experiment of Lexical-Semantic Tagging of an Italian Corpus

Ornella Corazzari[1], Nicoletta Calzolari[2], Antonio Zampolli[2]

Consorzio Pisa Ricerche[1]
P.za A. D'Ancona, 1 - 56100 Pisa
Istituto di Linguistica Computazionale (ILC)[2]
Area della Ricerca del CNR –San Cataldo
Via Alfieri 1 - 56010 Pisa
corazzar@ilc.pi.cnr.it
glottolo@ilc.pi.cnr.it.
pisa@ilc.pi.cnr.it

Abstract

The availability of semantically tagged corpora is becoming a very important and urgent need for training and evaluation within a large number of applications but also they are the natural application and accompaniment of semantic lexicons of which they constitute both a useful testbed to evaluate their adequacy and a repository of corpus examples for the attested senses. It is therefore essential that sound criteria are defined for their construction and a specific methodology is set up for the treatment of various semantic phenomena relevant to this level of description.

In this paper we present some observations and results concerning an experiment of manual lexical-semantic tagging of a small Italian corpus performed within the framework of the ELSNET project. The ELSNET experimental project has to be considered as a feasibility study. It is part of a preparatory and training phase, started with the Romanseval/Senseval experiment (Calzolari et al., 1998), and ending up with the lexical-semantic annotation of larger quantities of semantically annotated texts such as the syntactic-semantic Treebank which is going to be annotated within an Italian National Project (SI-TAL). Indeed, the results of the ELSNET experiment have been of utmost importance for the definition of the technical guidelines for the lexical-semantic level of description of the Treebank.

Introduction

In this paper we present some observations and results concerning an experiment of manual lexical-semantic tagging of a small Italian corpus performed within the framework of the ELSNET project. The ELSNET experimental project has to be considered as a feasibility study. It is part of a preparatory and training phase, started with the Romanseval/Senseval experiment (Calzolari et al., 1998).

Given the rather small size of the ELSNET corpus to be annotated, in the preparatory phase we decided to concentrate the lexical-semantic annotation on the predicate-argument part of the sentences, which can be considered the core of a sentence and is crucial for semantic interpretation.

The ELSNET corpus is composed of 1000 contexts of 20 selected verbs (50 contexts for each verb) extracted from the journalistic section of the Italian PAROLE corpus (1996-1998).

Such a corpus type, composed of a significant set of semantically annotated examples of verb senses and their argument heads, allows us:

- to study the disambiguation task with respect to the verbal head and its arguments (to what extent the disambiguation of one of them has an impact on the disambiguation of the other elements);
- to analyse different aspects of verbs (e.g. the possibility to draw a list of typical semantic subjects vs. direct objects of a verb sense; the influence of the verb sense on the meaning of the subjects and direct objects which combine with it and viceversa; the adequacy of the used semantic types with respect to all the previous tasks, etc.).

In this paper we focus on:

- the methodology for lexical semantic tagging and the strategies for the treatment of some phenomena relevant to this level of annotation (such as titles, proper nouns, idioms etc.);
- some interesting aspects emerged from the analysis of the annotated verbs and their argument heads (e.g. the usefulness of using a lexicon enriched with semantic types).

Finally, few observations are provided about the limits of lexical-semantic annotation, in other words, about what cannot be expressed through lexical tagging.

1 A Brief Description of the Experiment

The ELSNET experiment was performed through different steps:

- **verbs selection**: verbs were supposed to represent different semantic fields (i.e. speech acts (*chiedere, chiamare*), mental verbs (*comprendere*), movement verbs (*entrare, portare*), perception verbs (*vedere*), etc.), and various subcategorization properties (transitive, intransitive, reflexive verbs, etc.);
- **corpus contexts selection**: contexts were supposed to illustrate the different meanings of the same verb, and display a significant variety of argument heads for each verb sense;
- **corpus annotation** at three different levels of description: morphosyntactic, functional, lexical-semantic.

At the lexical-semantic level, the corpus annotation was manually performed and consisted in both sense-tagging and semantic-tagging (Kokkinakis et al., 1999). By sense-tagging we mean the assignment, to corpus occurrences, of the appropriate sense taken from a lexical resource which is in our case the EuroWordNet (EWN) lexicon (Alonge et al., 1998). By semantic-tagging we mean the

assignment, to corpus occurrences, of the appropriate semantic type/concept (such as *human*, *animal*, etc.). In our case, semantic types are the ones defined within the SIMPLE project (Lenci et al., 1999).

The combined use of both the EWN lexicon and the SIMPLE ontology of semantic types, was decided in order to allow future comparisons of the two types of annotation and evaluation of the disambiguating power of the SIMPLE semantic types.

2 The Lexical-Semantic Annotation: the Treatment of some Problematic Cases

From the experiment of lexical-semantic annotation, it turns out that it is obviously of utmost importance to set up a strategy of annotation for some semantic phenomena such as idiomatic expressions, compounds etc., when a sense does not correspond to one single orthographic word. The ELSNET experiment was therefore useful to highlight issues which had to be considered and solved while defining the specifications for semantic tagging of a large corpus in the Italian National Project, where criteria are given for idioms, compounds, figurative uses, evaluative suffixation, proper nouns, foreign words, titles, etc. (SI-TAL, 2000).

Some of these phenomena are listed with more details in the following sections.

2.1 Compounds

Compounds were treated as a single unit and the internal components were linked through an underscore as shown below. This treatment is justified from a linguistic point of view because in most of the cases they are not semantically compositional or they are only partially compositional.

- *un_filo_di_continuita'*
- *professore_d'orchestra*
- *compagnia_di_prosa*
- *ombrello_antimissile*
- *alta_moda*

2.2 Proper Nouns

Proper nouns composed by at least two lexical items were treated as one entry as shown below.

- *Incisa_della_Rocchetta (proper noun)*
- *Pippo_Baudo (proper noun)*
- *Teatro_Stabile_delle_Erbe* (theatrical company/ troupe)
- *Amici_ della_farsa (theatrical company/ troupe)*

2.3 Titles

Titles composed by more than one lexical item are compositional sequences and the single components could be annotated at the semantic level, if we wanted to allow e.g. IR queries not only on the titles as such but also on the internal components of the title. However, in this experimental phase, titles were marked as single units in order to simplify the annotation strategy. It is worth noting that in the SI-TAL project, they will be annotated both at the level of the single components and as a unique sequence.

All titles are identified by a specific tag (stype=title). Their identification at the lexical-semantic level is desirable at least for the following reasons:

- for linguistic acquisition purposes, in order to obtain more coherent data (e.g. considering the sequence *pubblicare (to publish) 'I fiori del male'*, if titles were not annotated we could draw the wrong conclusion that one of the typical objects of *pubblicare* is not a book/title/semiotic artifact, but a flower/natural kind);
- for MT purposes, in order to translate correctly titles which frequently have no literal/equivalent translation or are left in their original language.

Few corpus examples follow:
- *Ditegli_ sempre_ di_ si (title of a show)*
- *Si_recita_Feydeau (title of a show)*
- *La_Corrida (title of a show)*

2.4 Figurative Uses and Idiomatic Expressions

Figurative uses and idiomatic expressions in general are marked with specific features. Their identification is important at least for the following reasons:

1. for MT purposes, since in many cases they have no exact lexical and, as far as idioms are concerned, structural equivalents;
2. for linguistic acquisition purposes, in order to obtain a correct data extraction (e.g. considering the sentence *non comprendo la **molla** di una simile violenza* (I don't understand the reason of such a violence), the extraction of the direct objects of the verb *comprendere* (to understand) would lead to the wrong conclusion that one of its typical objects is an 'artifact' (*molla*/ spring) of type 'product' (some artifacts indeed can be used in this position: *non comprendo i suoi dipinti/libri* (artwork/ semiotic_artifact));
3. for lexicographic purposes, in order to extend existing computational lexicons with new idioms, collocations, and lexicalized metaphors, and allow studies on them.

2.4.1 Metaphors

The following are examples taken from the ELSNET corpus:

- *Gli episodi di Lecce lasciano sgomenti anche perchè, risulta difficile comprendere la **molla** di una violenza di queste proporzioni.* (The events of Lecce dismay also because it is difficult to understand the reason of such a violence)
- *abbandonare la **passerella** dell'alta moda* (to abandon the haute couture)
- *lo Stato stesso che la cultura tradizionale del Mezzogiorno percepisce come un ingiusto **patrigno**...* (the government which is perceived by the traditional culture of the South as an unfair stepfather)
- *questo tenore che ... è arrivato fino alle **vette*** (this tenor which... is arrived till the top)
- *abbandonare la **strada** dello sport* (lit. trans.: to abandon the road of the sport)

The figurative uses are marked with a specific feature (fig=fig). The distinction between lexicalized and non lexicalized methaphors was ignored in this experimental

project, while it will be taken into account within the SI-TAL project.

2.4.2 Metonymy

Metonymy, which raises the same problems of data intepretation as the other figurative uses in general, are marked by a specific feature (fig=meton). For example, in the following corpus context,

- *una banda di ragazzi che tagliavano le* **gomme** (a group of children which were punching pneumatics)

gomma is annotated as 'template_type=material' and 'fig=meton'.

2.4.3 Idiomatic Expressions

Idiomatic expressions such as the following ones were treated as a single unit and a specific feature (fig=idiom) was assigned. E.g.:

- *il processo* **entra nel vivo** (to enter into the heart of the process)
- **entra in scena** *il monitoraggio* (the monitoring comes into play)
- **aprire un** *nuovo* **capitolo** *nell'industria* (lit. trans.: to open a new chapter in the industry)
- **tagliare la testa al toro** (lit.trans.: to cut the head of the bull)

3 Some Remarks about the Annotated Verbs and Argument Heads

With the availability of a relatively small semantically annotated corpus at the lexical level, it is not possible to draw strong conclusions, but some observations can be done about issues such as the possibility to determine typical subjects or direct objects combining with a given verbal sense, the usefulness of using a lexicon enriched with semantic types and/or collocations, criteria for disambiguating senses, etc.

3.1 Typical Semantic Arguments of a Verb

From the analysis of the semantically annotated corpus, it turns out that there are various ways of describing in terms of semantic types a typical argument of a given verbal sense. The arguments combining with a verbal head can be:

- semantically restricted: in this case, it is possible to define the specific semantic types which combine with it (**selection restrictions**);
- semantically completely unrestricted (**no selection restrictions**);
- semantically unrestricted but it is possible to define which semantic types cannot combine with it for sure (indeed, this is particularly important when a given semantic type allows us to discriminate between different senses of the same verb) (we could call it a **negative restriction**);
- partially semantically restricted: a list of preferences in terms of semantic types can be defined (**selection preferences**).

Let us consider as illustrative example the verb *arrestare*:
1. The first meaning of the verb means 'to stop'. According to our tagged corpus its typical arguments are the following:
subj= act; cause_act; natural_substance; purpose_act; time

dobj= non_relational_act; change_of_value; mouvement_of_thought; event; act; cause_act; cause_natural_transition

In many cases the direct object has a **negative connotation**.

Summing-up, the sense 'to stop' selects an almost unrestricted subject and direct object. However the subject is preferably non-human (indeed it can be also human, e.g. *il governo ha arrestato l'inflazione/ the government stopped inflation*) while the direct object is preferably an 'event; act; change; phenomenon...', but it seems that it cannot be for sure a human or human-like (human-group, institution, etc.) semantic type. Moreover the direct object has preferably a negative connotation. All this can be broadly expressed in the following way:

SUBJ:
preference= non-human;
DOBJ:
preference= event; act; change; phenomenon
preference= negative connotation
negative_restriction= human

Table 1: arg.s description

2. The second meaning of the verb is 'to arrest'. The arguments are:
subj= human; human_group; institution; profession
domain=military; law
dobj= human; agent_of_temporary_activity; agent_of_persistent_activity; kinship; profession; people

In many cases the direct object has a **negative connotation**.

Summing-up, the sense 'to arrest' clearly selects a human or human-like subject and direct object. The subject preferably belongs to the military/law domain, while the direct object preferably has a negative connotation (but not always, for instance *arrestare un innocente/ to arrest an innocent*). This is shown in the following table:

SUBJ:
selection_restriction= human or human-like
preference= institution, human_group, profession
preference= domain=law, military
DOBJ:
selection_restriction= human
preference= negative connotation

Table 2: arg.s description

Another example is the verb *percepire*. The direct objects of this verb can be described as follows:
1. The first meaning of the verb is 'to perceive' with the senses. This sense is marked as a **'perception'** verb.
dobj= color; group; shape; sign; phenomenon
2. The second sense is 'to receive' and is marked as **'change_possession'**
dobj= money; convention; number; amount
3. The third sense is a figurative use of the verb (to perceive with the intuition) (to perceive something as if it is something else) and is marked as **'perception figurative'**
dobj= unrestricted

It must be said that this last sense frequently occurs in our corpus with a modifier introduced by *come* (ex: *l'opinione pubblica *percepisce* il Servizio sanitario nazionale (Ssn) come poco efficiente/ public opinion perceives the Ssn as not so efficient*) but not necessarily.

Summing-up, for *percepire* only the second sense seems to have a semantically restricted direct object. At last, the third meaning is marked in many cases by a specific (preferred) syntactic pattern.

3.2 Verb/Arguments Interaction at the Lexical-Semantic Level

The interpretation of the sense of a given argument head may strongly depend on the meaning of the surrounding context, more precisely, of the verbal head. Between the verb and its argument heads there is a strong interaction from the semantic point of view: the verb meaning may determine (or select) the sense of its subject and/or direct object. For instance *arrestare* (to arrest/ to stop), as said above, frequently selects direct objects which have themselves or receive from the verb a negative connotation in most of the cases, as shown below:

Dobj	Sem.type of Dobj	Conn. Feat.
ladro_1	agent_temp_act	neg
spacciatore_1	agent_temp_act	neg
trafficante_1	agent_temp_act	neg
traffico_2	act	neg
invasione_1	cause_act	neg
massacro_1	cause_nat_trans	neg
inflazione_1	event	neg
pregiudicato_1	human	neg
balordo_1	human	neg
maniaco_1	human	neg
strozzino_1	agent_temp_act	neg

Table 3: Dobj of the verb *arrestare*

Another example is the verb *comprendere* which with the meaning of 'to include' selects a specific sense of certain lexical items when they are subjects of the verb. For instance the lemmas below are marked in the SIMPLE lexicon as 'group of more or less specified entities', which can 'include' other entities:

Dobj	Sem.type of Dobj
carico_1	group
elenco_1	group
equipaggiamento_2	group
lista_2	group
panorama_1	group
tris_1	group
comune_1	human_group
consiglio_2	human_group
costituente_2	human_group
dossier_1	group

Table 4: Dobj of the verb *comprendere*

It may also happen that the sense of the direct object determines the meaning of the verb. For instance, the semantic type of the direct object helps to characterize the different possible senses of the verb *coprire*, as shown in the table below:

- *coprire un periodo* (to cover a period of time):

Dobj	Sem.type of Dobj
spazio(no)	time
1970-1993	time

Table 5: Sem.Type of Dobj

- *coprire uno spazio* (to cover a space/ a distance):

Dobj	Sem.type of Dobj
superficie_1	area
territorio_1	area
area_2	area
area_1	area
pista_1	artifactual_area
continente_1	geopolitical_location
80_per_cento	part
35%	part

Table 6: Sem.Type of Dobj

- *coprire una persona/un reato* (to hide a crime) :

Dobj	Sem.type Of Dobj	Conn. Feat.
crimine_1	act	neg
mafioso_2	human	neg
violento_1	human	neg

Table 7: Sem. type of Dobj

- *coprire un suono* (to smother):

Dobj	Sem.type of Dobj
rumore_1	experience_sound

Table 8: Sem. Type od Dobj

3.3 Acquisition of Senses and Enhancement of Existing Lexical Resources

The analysis of a semantically tagged corpus allows not only to identify totally new senses but also to have a more precise and complete view on the semantics of a lemma and to decide on a more sound base than human intuition which and how many senses to encode for the same lemma. Relying on the different semantic types of argument heads that combine with a given verb, it is possible to identify the most general senses of a lemma, to capture the most specific senses or shifts of meaning of the same lemma, and to decide to collapse some uses into more general, inclusive senses according to:

- different needs and requirements of the lexical resource to be created/extended/tuned (indeed, both the number and the type of senses to be encoded may strongly depend on the 'apparatus' (information types) used for describing them, e.g. semantic nets, frames and selection restrictions, type of ontologies, domain information, semantic relations, etc.);
- different applications of the lexical resource (e.g. MT, IR, etc.). For instance, in an MT environment (bilingual, multilingual resources), it makes sense to treat as independent meanings the ones that have a different translation (e.g. the sense number 9 below, among others, for the language pair Italian/English).

Therefore corpus analysis does not necessarily leads to an excessive sense distinction which is not desirable for different reasons (Calzolari et al., forthcoming; Fellbaum (ed.), 1998), but may provide ground for decisions based on actual evidence.

We provide below the example of the verb *abbandonare* (to abandon/leave) which has at least the following three main senses according to current paper dictionaries:
- to abandon, to leave forever (e.g. a place)
- to abandon, to desert (e.g. the children)
- to give up, to renounce

On the basis of the analysis of the semantic types of direct objects, the following major/minor senses (uses) of *abbandonare* come out:

1. 'to leave a place': **dobj**= building, geopolitical_location, area ..
2. 'to get up' (*abbandonare la sedia, un veicolo*): **dobj**= furniture, vehicle
3. 'to abandon someone': **dobj**= kinship, animal, human
4. 'to give up an activity': **dobj**= act, purpose_act
5. 'to give up an ideology, a dream..': **dobj**= movement_of_thought, cognitive_fact
6. 'to leave a group, a party, a club..': **dobj**= institution, human_group
7. 'to abandon a sector, a domain…(sport, biology)': **dobj**= domain
8. 'to change one's psychological state' (*abbandonare la calma, la prudenza*/ lit. trans.: *to abandon the calm, the caution*): **dobj**= psych_property
9. 'to drop something' (*abbandonò la divisa a casaccio sulla sedia*/ *he dropped the uniform at random on the chair*): in this case the direct object is a concrete/ inanimate entity (neither human nor animal) which is in the corpus example a 'clothing'. It is worth noting that this specific use of *abbandonare* in the corpus example combines with a particular modifier which cannot occur with the other senses of the verb (e.g. **abbandona la moglie a casaccio*/ **he abandoned the wife at random*).

4 The Complexity of Word Sense

Word sense disambiguation can be performed through a combined interpretation of different levels of information: morphosyntactic/ syntactic/ semantic and even multilingual (Gale et al., 1992). For instance, other projects, such as DELIS (Monachini et al., 1994), stressed the deep interaction between e.g. morphosyntactic patterns and word meanings. The following are syntactic and semantic indicators which can help sometimes in the identification of a sense, but they are not at all sure tests: they have only a partial application or they are not completely discriminating. Therefore human judgement has still an important part in word sense disambiguation.

- a specific syntactic pattern allows to select a particular sense. This is the case of *comprendere* which can co-occur with a that-clause when it means 'to understand' (and not when it means 'to include'), or *aprire* which occurs in the pattern 'to open to a human' when it acquires the meaning of 'to be ready, willing, open, well disposed towards someone' (e.g. *Cossiga apre a La Malfa*) (Monachini et al., 1994; Calzolari et al., 1996; Atkins et al., 1988);
- the semantic domain of use of a given lemma can help to select a specific word meaning (e.g. *perseguire un reato/ to prosecute a crime* (domain=law));
- a specific modifier sometimes selects or preferably occurs with a particular sense. For instance *perseguire* **penalmente**/ *to prosecute at the penal level* does not mean for sure 'to pursue (a goal) '; *comprendere* **benissimo**/ *to understand very well* does not mean for sure 'to include' (Monachini et al., 1994; Calzolari et al., 1996);
- a specific class of subjects and/or direct objects and/or indirect objects etc. can help to select a particular meaning of a word (e.g. the human subject always selects the meaning 'to understand' of the verb *comprendere*);
- different synonyms and/or antonyms select different senses (Cruse, 1986);
- two different senses of a lemma cannot be selected simultaneously by the same context (Cruse, 1986) (e.g. **Leo arresta sia il colpevole che il corso degli eventi/ *Leo arrests both the criminal and the events*);

It is clear that the availability of large quantities of semantically tagged corpora may help in better analyse the impact of different clues for word sense disambiguation and the interaction of clues at different levels of linguistic description.

5 What Cannot be Easily Encoded at the Lexical-Semantic Level of Annotation

In a large number of cases, sense interpretation requires appeal to extralinguistic knowledge (world knowledge, etc.) which cannot be encoded or, to put it that way, captured at the lexical-semantic level of description. We provide below a few examples:

1) When the metaphors are not restricted to a single lemma (e.g. *la* **chiave** *del problema/ the clue to a problem*) but extended to an entire sequence. E.g.:
   - *l'auto verde arriva sul tavolo del governo* (lit.trans.: *the green car arrives on the table of the government*)

The sequence means that the 'topic' of *auto verde* (the car which does not pollute) will be discussed by the government. However, at the lexical level only *auto*, *verde* and *tavolo* will be marked as figurative uses. Whereas the general unusual sense will come out only from a violation of the selection restrictions. Indeed, a car (template_type=vehicle) can arrive (template_type=move) but not on a table (template_type=furniture), more probably in a place/location, while the 'topic' of the *auto verde* can arrive on the table of the government. This complex

sense interpretation cannot be expressed through lexical-semantic annotation. It is impossible to immagine the assignment of a label 'topic' to *auto* and/or *verde* (everything can be indeed a topic).

2) When a sequence is ambiguous between two meanings according to the intention of the author:

- *Titolo: Nina Vinchi **entra in scena** (Title: Nina Vinchi starts/ comes on stage)*
  Sottotitolo: A 84 anni la signora del Piccolo affronta per 3 ore i giudici.

In this corpus context the expression *entrare in scena* has the double meaning of to appear/to start (the idiomatic sense) and to come on stage (the literal sense). In this case, the interpretation of the sequence is based on knowledge about the domain type (domain=theater) and the context type (indeed ambiguities of this kind are frequently used within titles).

3) When some words acquire a specific sense strictly dependent on the context in which they occur, which cannot be encoded at the lexical-semantic level. E.g.:

- *la donna (Pauline Collins), che ha gia' visto arrestare il marito dai **tedeschi**, viene fermata*

Arrestare usually combines with a subject belonging to the military/law domain. Also in this case *tedesco* has to be interpreted as 'German soldier' (and not any kind of German people). However, in the computational lexicon *tedesco* is obviously marked as 'people' and cannot be otherwise.

4) Another example similar to the previous one is the verb *chiamare* which, sometimes, means (is synonymous of) to telephone. The identification of this sense strongly depends on a complex process of context interpretation in most of the cases (indeed there are few cases in which the interpretation is easy, for instance when the direct object is not a human but a phone number or an inanimate entity, e.g. *'chiamare il (numero) 113/ Buckingham Palace/ l'ambulanza'*). E.g.:

- *E io **chiamo** Craxi per 150 miserabili milioni?(And I should call Craxi for 150 miserable milions?)*

- *Il giorno dopo Finocchi **chiamo'** Volpari: "... Purtroppo, non ho nessuna notizia (The day after Finocchi called. Volpari: "...Unfortunately, I have no news)*

- *In gran parte sono bambine dai 6 ai 14 anni. **Chiamano** per lo piu' da Milano e provincia, di preferenza al mattino (In most of the cases they are girls from 6 to 14 years. They call in most of the cases from Milan and surrounding areas, preferably in the morning)*

5) At last, we provide the example of *tagliare/ to cut*. From the table below it is evident the complexity and variety of senses and actions implied by the verb according to the kind of direct object which combines with it. Not all these shifts of meanings can/must be captured through lexical-semantic annotation (sense and semantic tagging). For instance, *tagliare il prato* (to cut the grass) means to eliminate/ reduce the grass; *tagliare le gomme* means to make a hole (to punch) in the pneumatic; *tagliare i capelli* means also to give a nice shape to the hair (not necessarily to shorten them); *tagliare il mantello* means 'to divide' the mantle; *tagliare la legna* means to cut into pieces the wood; *tagliare le corolle* means to detach the corolla of flowers (to separate the corolla from the flower); on the other hand, *tagliare una fettina* (to cut a small slice) moves the focus from the cut entity to the cut part, etc.

Dobj	Sem.type of Dobj	Sem.type of the Verbal Head
prato	area	cause_change_of_state
gomma	artifact	cause_change_of_state
stoffa	artifactual material	cause_constitutive_change
lingua	body_part	cause_constitutive_change
testa	body_part	cause_constitutive_change
capello	body_part	cause_change_of_state
mano	body_part	cause_constitutive_change
mantello	clothing	cause_constitutive_change
legna	material	cause_constitutive_change
corolla	part	cause_constitutive_change
fettina	part	cause_constitutive_change
pezzo	part	cause_constitutive_change
cespuglio	plant	cause_constitutive_change
spino	plant	cause_constitutive_change

Table 9: Dobj of *tagliare*

Conclusion

The ELSNET experiment allowed us to better understand some of the problematic aspects of lexical-semantic corpus annotation and to have a broad overview of some of the possible type of analysis which can be done on a corpus tagged at the lexical semantic level.

Semantically tagged corpora larger than the one of ELSNET can be useful, for instance:

- to evaluate the disambiguating power of the semantic types of a given lexical resource used for lexical-semantic corpus annotation;
- to assess the need of integrating traditional dictionaries or computational lexicons with senses attested in the corpus;
- to identify the inadequacy of certain sense distinctions attested in traditional dictionaries or current computational lexicons which are not applicable (see Calzolari et al., forthcoming);
- to check the real frequency of already known senses (some of them can be scarcely attested, or the reverse, in a specific corpus type (e.g. journalistic corpus)).

References

Atkins, B.T., Kegl, J., Levin, B. (1988). Anatomy of a Verb Entry: from Linguistic Theory to Lexicographic Practice. International Journal of Lexicography, 1, 84--126.

Alonge, A., Calzolari, N., Vossen, P., Loksma, L., Castellon, I., Marti, M.A., Peters, W. (1998). The Linguistic Design of the EuroWordNet Database, Special Issue on EuroWordNet. Computers and the Humanities, 32 (2-3).

Busa, F., Calzolari, N., Lenci, A., Pustejovski, J. (1999). Building a Lexicon: Structuring and Generating Concepts. In Proceedings of the Computational Semantics Workshop. Tilburg.

Calzolari, N., Corazzari, O. (forthcoming). Senseval/Romanseval: the framework for Italian. Computers and the Humanities. Dordrecht.

Calzolari, N., Corazzari, O., Monachini, M., Roventini, A. (1996). Speech Act and Perception Verbs: Generalizations and Contrastive Aspects. In EURALEX-96 Proceedings (pp. 73--83). Goteborg University.

Corazzari, O. (1992). Phraseological Units. NERC--92--WP8--68. ILC, Pisa.

Cruse, D.A. (1986), Lexical Semantics. Cambridge, Cambridge University Press.

Fass, D. (1991). met*: A Method for Discriminating Metonymy and Metaphor by Computer. Computational Linguistics, 17(1), 49--90.

Fellbaum, C. (ed.) (1998). Wordnet, An Electronic Lexical Database. Cambridge, MIT Press.

Gale, A. W., Church, K.W., Yarowsky, D. (1992). A Method for Disambiguating Word Senses in a Large Corpus. Computers and the Humanities, 26, 415--439.

Kilgarriff, A. (1993). Dictionary word sense distinctions: An enquiry into their nature. Computers and the Humanities, 26, 365--387.

Kokkinakis, D., Kokkinakis, S. J. (1999). Sense-Tagging at the Cycle-Level Using GLDB. Goteborgs Univerity.

Lenci, A., Busa, F., Ruimy, N., Gola, E., Monachini, M., Calzolari, N., Zampolli, A. (1999). Linguistic Specifications, SIMPLE Work Package 2, Deliverable D2.1. University of Pisa and ILC of CNR.

Monachini, M., Roventini, A., Alonge, A., Calzolari, N., Corazzari, O. (1994). Linguistic Analysis of Italian Perception and Speech Act Verbs. DELIS Working Paper. ILC, Pisa.

Montemagni, S. (1990). Definition, General Features and Taxonomy of Multi Word Expressions. ILC, Pisa.

PAROLE (1996-1998). Preparatory Action for Linguistic Resources Organization for Language Engineering. LE-4017, Language Engineering.

Rodriguez, H., Climent, S., Vossen, P., Loksma, L., Peters, W., Alonge, A., Bertagna, F., Roventini, A. (1998). The Top-Down Strategy for building EuroWordNet: Vocabulary Coverage, Base Concepts and Top Ontology. Special Issue on EuroWordNet. Computers and the Humanities, 32 (2-3).

SI-TAL (2000). Specifiche Tecniche di SI-TAL. Manuale Operativo.