

Encoding data on lexicalization of semantic components: the EuroWordNet 'relational' model

Abstract

In this paper we discuss the approach being followed to encode data on lexicalization patterns of semantic components in verb roots, within the framework of the EuroWordNet project. We focus on how the relational model of the EuroWordNet database is being 'used' to encode information on the semantics of Italian verbs. First, we provide some general information on the linguistic design of the semantic database we are building. Then, the kind of information which is being encoded for Italian verbs is discussed in detail, by taking into consideration semantic/syntactic verb classes. Furthermore, we also refer to theoretical and computational uses of the data on lexicalization patterns we are encoding.

Keywords: Lexicalization, semantic component, EuroWordNet, lexical-semantic relation

1. Introduction

Research on lexical semantics has recently become a major concern both for the theoretical field of linguistic studies and for computational applications. A trend has emerged towards integrating syntactic treatments with descriptions of semantic properties of words within lexical resources developed mainly for Natural Language Processing (NLP) applications but also for human use. At the same time, the need for data on the semantics of words has constantly increased as new NLP systems have been developing which have to deal with the major problem of automatic sense disambiguation.

Although the Princeton WordNet database (Miller et al. 1993) was not born to fulfill the need for semantic data by computational applications, it has in fact become the most used semantic database for research in many areas of computational linguistics. Its relational design, inspired by psycholinguistic theories of human lexical memory, provides data on the semantic contents of words which can be advantageous in general for tasks connected with Word Sense Disambiguation (e.g., Agirre and Rigau 1996; Sanfilippo 1997; etc.) and, in particular, for Information Retrieval systems (e.g., Bloksma et al. 1996; Richardson and Smeaton 1995; etc.).

However, the Princeton Wordnet is a resource containing data only for English while many applications now need data to work in a cross-linguistic perspective. The main goal of the EuroWordNet (EWN) project¹ is thus to develop a multilingual lexical resource, retaining the basic underlying design of Wordnet (in particular, of the database version WordNet 1.5, hereafter WN 1.5) while at the same time trying to *improve* it in order to answer the needs of research in the computational field. Of course, in the process of building the semantic database, we are also taking into account hypotheses recently put forward in the theoretical field of studies in lexical semantics in order to obtain a tool which is rich enough to be used for the most varied tasks. Thus, the set of lexical relations to be encoded between word meanings has been extended or modified in various ways with respect to that encoded in WN 1.5 and relations applying between words in the languages involved have also been added.

The main goal of this paper is to discuss how the EWN relational model is being used to encode information on lexicalization of semantic components within the roots of Italian verbs. We shall also briefly indicate the possible 'practical' usage of the data we are encoding. The paper is organized as follows: in section 2 a brief overview of the characteristics of the EWN relational model is given, with emphasis on newly identified relations which are useful for encoding data on lexicalization patterns. In section 3 examples of the kinds of data treated are provided, by dealing with some Italian verbs. The theoretical and practical relevance of the results of our work are also briefly indicated. In section 4 some concluding remarks are added.

2. The EuroWordNet model

2.1. A relational linguistic approach

The avowed goal of the Princeton Wordnet was that of instantiating hypotheses based on results of psycholinguistic research (Miller et al. 1993). Thus, the database is organized as a semantic network in which the meanings of words are represented in terms of their conceptual-semantic and lexical relations to other words. Moreover, the structures imposed on the noun, verb and adjective networks are somehow different to reflect a different way of categorizing experience.

Although the basic structure of the EWN database is shaped on that of WN 1.5, some important changes have been made to that overall design due to various theoretical and practical reasons. First of all, though we do not want to neglect the importance of taking into account results on the psycholinguistic research side, we aimed at assuming a purely 'linguistic' perspective in our work. Since WN 1.5 is organized as a conceptual ontology, artificial levels were introduced for concepts which are not lexicalized in English in order to achieve a more compact and coherent structure reflecting how lexical knowledge and memory are organized. In EWN, instead, we are trying to build a purely linguistic ontology by encoding only data on lexicalized meanings and semantic relations occurring between words in each of the languages involved. As EWN is a multilingual resource, it is important for us to be able to encode similarities but also differences in lexicalization patterns of semantic components among the languages under analysis. A further, more practical, reason for the changes made is that we use machine-readable dictionaries as main sources of data: these are rich sources of semantic information but are structured in such a way that we need partially different relations with respect to those encoded in WN 1.5 in order to encode significant semantic data.

The point of view we have assumed is very similar to Cruse's approach to the meaning of a word (Cruse 1986). In his view the meaning of a word is conceived of as a kind of 'semantic field', containing all of the possible (grammatical) sentential contexts of the word and all of the possible (grammatical) substitutes within those contexts. Thus, the meaning of a word is made up, at least in part, of (the meanings of) other words which therefore correspond to 'semantic components' of the first word. Such semantic components carry the lightest possible burden of theory and, in fact, no claim is made that they are either primitive, functionally discrete, universal or drawn from a finite inventory. Furthermore, in Cruse's view there is no claim that the meaning of any word can be exhaustively characterized by any finite set of such semantic traits. In keeping with Cruse, we assume a relational view of the

lexicon according to which all the semantic aspects pertaining to the lexical level are reflected in the paradigmatic and syntagmatic relations occurring between any two words in a language. Therefore, the meaning of a word can be 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. The adoption of this approach to word meaning has, however, some consequences already pointed out by Cruse (1986:19): firstly, that drawing a line between the meaning of a word and 'encyclopaedic' facts concerning the extra-linguistic referents of the word is not really possible; then, that we are not actually interested in isolating 'pragmatic meaning' as a separate domain of lexical meaning. Indeed, we are trying to capture (language-dependent) lexicalization patterns of semantic components, wherever this is possible, without claiming that we are comprehensively encoding all possible semantic references, but also without drawing a sharp distinction between what is strictly speaking 'semantic' and what could be described as 'pragmatic meaning'.

2.2. Relations in EWN

As in WN 1.5, the basic notion around which the whole EWN database is being built is that of a *synset*, i.e. a set of synonymous words.² Then, whereas the main relations encoded in WN 1.5 are hyponymy and meronymy for nouns, and troponymy³, entailment and cause for verbs, in EWN also additional relations have been defined, and various other changes have been made to the WN 1.5 overall organization of relations.

First of all, some labels have been identified which can be added to the relations (i.e. conjunction or disjunction of relations, negation, factivity and intention of causal relation), making the implication of the relations more explicit and precise. Secondly, whereas in WN 1.5 a rigid distinction is drawn among different parts-of-speech (PoSs) and each PoS forms a separate system of language-internal relations, in EWN relations applying between different PoSs have been added. Finally, some relations already present in the WN 1.5 design have been modified in order to specify their role more clearly.

In this paper, we are not going into the details of each of the relations being encoded in EWN (discussed instead in Alonge 1996; Climent et al. 1996; and Alonge et al. forthcoming). In what follows, we shall rather discuss issues connected with the main relation which is being used to identify lexicalization patterns of semantic components in verb roots: the *Involved* relation and its subtypes.

2.3. Encoding information on verb meanings

Both practical and theoretical reasons led us to add a so-called *Involved* relation to our set of links. This relation was, first of all, needed to differentiate shallow hierarchies occurring in the networks derived from our sources. In fact, many classes of words are defined by means of a phrase whose syntactic head (the hyperonym or *genus term*) is not very 'significant' since it refers to a general concept which also applies to many other words displaying, however, significant meaning differences. What follows it (the *differentia*) often provides a much better characterization of the word meaning. This can be seen by taking into account some hyponyms of the Italian verb *andare*⁴, derived from the Italian Lexical Database:

<i>camminare</i> to walk	=	“andare a piedi” to go <i>on foot</i>
<i>pattinare</i> to skate	=	“andare sui pattini” to go <i>on skates</i>
<i>navigare</i> to sail	=	“andare (detto di navi)” to go (<i>said of ships</i>)
<i>uscire</i> to go out, to exit	=	“andare fuori” to go <i>out</i>
<i>entrare</i> to go in, to enter	=	“andare dentro” to go in
<i>coricarsi</i>	=	“andare a letto” to go <i>to bed</i>

While *andare* (to go) simply indicates *motion along a path*, its hyponyms may refer to very different ‘kinds of motion’:

- *camminare* is an *undirected manner-of-motion* verb;
- *pattinare* refers to *undirected motion by means of a vehicle*;
- *navigare* refers to *undirected motion performed (only) by specific vehicles*;
- *uscire* and *entrare* refer to *motion from and to partially-specified locations*;
- *coricarsi* indicates *motion to a specific place* (incorporated within the meaning of the verb itself).

If we only coded an IS-A or hyponymy relation with *andare* for these verbs, we would lose much important semantic information on them. Thus, we defined the *Involved* relation to encode a link between a verb (or also a noun referring to a state, process or event) and a concrete noun (or, in some cases, an adjective or an adverb)⁵ whose meaning is incorporated in (or strongly connected with) the verb (or noun) meaning.⁶ In addition to the general underspecified relation (used for unclear cases of ‘involvement’), the specific relation subtypes indicated in Table 1 can be encoded⁷:

Table 1

INVOLVED_AGENT the active causer of the process/event referred to by a verb or noun	sgambettare/neonato (to kick one’s (small) legs about) /(baby)
INVOLVED_PATIENT the ‘entity’ undergoing, or being affected by, a process/event/state	to teach/learner
INVOLVED_INSTRUMENT any object used to do something	to paint/paint-brush
INVOLVED_LOCATION the location where a situation takes place	to swim/water

INVOLVED_SOURCE_DIRECTION the source of a 'motion situation'	to disembark/ship
INVOLVED_TARGET_DIRECTION the target of a 'motion situation'	rincasare/casa (to go back home)/(home)
INVOLVED_RESULT the result of a change	freeze/ice
INVOLVED_MANNER the manner in which a situation takes place	shout/loudly

The *Involved* relation is only being used to encode data on arguments/adjuncts that are *strongly* implied in the meaning of a verb/noun (and generally indicated within dictionary definitions). This is not the same as encoding the arguments or adjuncts co-occurring with a verb/noun in a sentence. In the relational approach we follow, we only encode the semantic features *incorporated* in the meaning of a word. These certainly also determine the kind of syntactic contexts in which that word may occur, but do not necessarily coincide with them. For instance, whereas a verb like *to move* allows agent arguments, there is no inherent reference to a particular 'involved-agent' in its meaning (because, in fact, many kinds of 'agents' can move). No INVOLVED_AGENT relation is therefore encoded for *move*. However, the Italian verb *sgambettare*, meaning *to kick one's (small) legs about* and referring to a movement typically performed by babies, clearly incorporates the 'agent-protagonist' *baby*. This information can therefore be encoded by means of the relation INVOLVED_AGENT. Thus, the various relations are only being encoded when there is a *clear* and *strong* association with another word in the meaning of a verb/noun, generally indicated in dictionary definitions.

The relevance of encoding this type of information is apparent in the most recent theoretical developments in lexical semantics. Research in this field has demonstrated that there is cross-linguistic variation with respect to the possibility of *conflating* certain semantic components within verb roots (Talmy, 1985). Furthermore, there is often a direct connection among the arguments/adjuncts lexicalized within a verb root (corresponding to some of its semantic components) and the verb's syntactic properties (cf. e.g., Jackendoff 1990, Levin 1993). Finally, the kind of semantic references lexicalized in a verb root correlate with the selectional preferences of the verb itself (e.g., Alonge 1994). This information is, thus, crucial not only for theoretical purposes but also for computational applications.

3. Encoding data on lexicalization patterns in Italian

In order to acquire the information which is necessary to encode lexical semantic data, we use as our main source the Italian Lexical Database, which was developed by merging data coming from different MRDs, but we also use textual corpora to individuate information missing within it. Furthermore, we are re-using data which were acquired within previous European research projects (e.g. *Acquilex* and *Delis*)⁸ by analysing both MRDs and corpora. However, the information found in the resources used is further elaborated by taking into account the results of recent theoretical research in lexical semantics.

Let's take into consideration the motion verbs cited above again, in order to clarify how we have encoded useful information on them by using EWN relations. As already seen, these verbs are all hyponyms of *andare*, which simply indicates *motion along a path* and is

underspecified with respect to the reference to a *direction* of motion. Indeed, this verb seems to function as a kind of *copula*: it may appear with either a directional complement or a non-directional one. In the former case, the VP refers to directed motion, in the latter it refers to undirected motion. By establishing a hyponymy relation with *andare*, however, we may only infer a reference to *motion along an unbounded path* for all the verbs in our example. This, however, is not sufficient to differentiate the various verb meanings. We have encoded more useful information by means of the Involved relation:

camminare (to walk)

HAS_HYPERONYM *andare*

INVOLVED_MANNER *a piedi* (on foot)

pattinare (to skate)

HAS_HYPERONYM *andare*

INVOLVED_INSTRUMENT *pattini* (skates)

INVOLVED_AGENT *pattinatore* (skater)

navigare (to sail)

HAS_HYPERONYM *andare*

INVOLVED_AGENT *nave* (ship)

entrare (to go in, to enter)

HAS_HYPERONYM *andare*

INVOLVED_SOURCE_DIRECTION *fuori* (outside)

INVOLVED_TARGET_DIRECTION *dentro* (inside)

uscire (to go out, to exit)

HAS_HYPERONYM *andare*

INVOLVED_SOURCE_DIRECTION *dentro* (inside)

INVOLVED_TARGET_DIRECTION *fuori* (outside)

coricarsi (to go to bed)

HAS_HYPERONYM *andare*

INVOLVED_TARGET_DIRECTION *letto* (bed).

By encoding these relations we may emphasize differences in lexicalization of semantic components within the roots of apparently corresponding verbs in different languages. For instance, whereas all English motion verbs may refer to change-of-position (Levin and Rappaport 1992) and may, therefore, occur with directional phrases, only a subset of Italian motion verbs may refer to directed motion (change-of-position), and can, thus, appear with directional phrases (Alonge 1994). In our database the inherent reference to directed motion can be encoded by using the INVOLVED_SOURCE_DIRECTION and/or INVOLVED_TARGET_DIRECTION relations. When neither of these relations is encoded for a verb, the verb does not refer to directed motion and cannot, therefore, appear with directional phrases (cf. **Gianni camminò a casa* vs. *Gianni walked home*; **Gianni pattinò a scuola* vs. *Gianni skated to school*; **La nave navigò al porto* vs. *The ship sailed to the harbor*). When, on the other hand, either of these relations is encoded, we may infer the possibility of a verb occurring with directional phrases (cf. *Gianni entrò nella stanza* (Gianni went into the room); *Gianni uscì dalla stanza* (Gianni went out of the room)).

Besides being useful to identify differences in lexicalization of semantic components, the information we have encoded may be used to obtain data on semantic co-occurrence restrictions. For instance, the goals implied within the meanings of *entrare* and *uscire* are ‘partially specified’, as indicated by the adverbs found in their definitions: in the former case there is a reference to a +INTERNAL place, in the latter to a +EXTERNAL place. Thus, only directional phrases displaying the appropriate semantic references will be allowed with the two verbs: e.g., *all’aria aperta* (in the open air) has to be excluded as a possible complement of *entrare*, since *aria aperta* does not occur in the taxonomy of ‘inner’ places.

When in the meaning of a verb there is a reference to a typical ‘specific’ argument, different co-occurrence restrictions may be inferred for that verb. *Coricarsi*, for instance, involves a reference to motion to a specific place (*letto* = bed). Thus, it cannot select *letto* as a complement (to avoid redundant information) unless this is somehow modified (cf. *Si coricò nel proprio letto* (He went to bed in his own bed)). On the other hand, the verb can occur with nouns referring to various objects which display some semantic/pragmatic similarity to *beds* (cf. *Si coricò su un giaciglio di paglia* (He went to bed on a straw bed); *Si coricò sul divano* (He went to bed on the sofa)). This can be due to the fact that the noun *letto* has two basic senses in Italian: i) the piece of furniture which is normally used to sleep on; ii) every piece of furniture or object where one can lie down. By establishing a relation with both these senses of *letto*, we may account for the various co-occurrence restrictions of the verb.

Pattinare also involves a reference to a specific argument, i.e. the instrument used to perform the action referred to by the verb. Thus, the verb cannot appear with a complement indicating such an instrument, unless additional information is provided (e.g., *Gianni pattinava con i pattini a rotelle* (Gianni was skating with roller-skates)). With respect to this verb, an INVOLVED_AGENT (*pattinatore* = skater) is also encoded. Since *pattinatore* is a hyponym of *persona* (person), we may infer that *pattinare* requires a human subject. This information can be useful to automatically disambiguate occurrences of different senses of *pattinare* in texts. In fact, *pattinare* has also another sense, defined as “slittare, detto di automobili” (to skid, said of cars). By taking into consideration the relations encoded for the two verb senses, it is possible to determine the appropriate sense of *pattinare* in sentences as, for instance: *Gianni pattinò sul ghiaccio* (Gianni skated on ice) and *La macchina pattinò sul ghiaccio* (The car skidded on the ice).

Also the hyponyms of the verb *mettere* (to put) display semantic characteristics which can be partly identified by means of the Involved relation. Their hyperonym ‘transfers’ to them the semantic reference to something being moved to a target by an involved agent. However, more specific semantic information is implied in their meanings, as can be seen in the following examples:

- archiviare* (to place in the archives) = “mettere in archivio un documento” (to place a document in the archives)
 INVOLVED_PATIENT *documento*
 INVOLVED_TARGET_DIRECTION *archivio*
- avvicinare* (to bring near) = “mettere una cosa vicino ad un’altra o a qlcu.” (to bring something near something or someone)
 INVOLVED_PATIENT *cosa*

INVOLVED_TARGET_DIRECTION *cosa*
INVOLVED_TARGET_DIRECTION *persona*
INVOLVED_TARGET_DIRECTION *animale*

caricare (to load) = “mettere un carico addosso a una persona o sopra un animale o una cosa” (to put a load or a burden on someone or on an animal or on something)

INVOLVED_PATIENT *carico*
INVOLVED_TARGET_DIRECTION *persona*
INVOLVED_TARGET_DIRECTION *animale*
INVOLVED_TARGET_DIRECTION *cosa*

From the relations encoded we may infer useful data on the syntactic constructions allowed by the verbs under analysis and on selectional preferences; moreover, we may infer information on differences in lexicalization patterns with respect to other languages.

Archiviare involves a patient which is partially specified in that it refers to a quite general concept which includes a number of more specific concepts. Therefore, all its hyponyms will be allowed as objects of the verb (cf., e.g. *Archiviò la pratica/l'atto* (He/She placed the dossier/the act in the archives)). Then, being the target inherently specified, no directional phrase will be generally allowed by the verb unless it provides additional information.

Avvicinare involves a partially-specified patient (a *thing*) and different types of targets. The information encoded with respect to the kind of patient involved can be useful to distinguish occurrences of different senses of *avvicinare*. Indeed, when appearing with human patients, *avvicinare* means *to approach someone*: cf. *Avvicinò le sedie* (He/she drew up the chairs) vs. *Avvicinò il ragazzo* (He/she approached the guy).

Finally, *caricare* involves different types of targets, indicated by means of a label *conjunctive* added to the relations, and thus allows different target-complements.

The examples discussed show that by using the EWN relations we may encode data on language-specific lexicalizations within verb roots. These data are directly linked to the syntactic properties of verbs and can be very useful for theoretical research. Work in the field of lexical semantics has demonstrated the existence of a strong connection between semantics and syntax (e.g. Levin 1993), however, the appropriate semantic components determining the syntactic behaviour of several verb classes (in different languages) still have to be identified. Thus, the information which is being encoded within the EWN database can be of great utility for research in this field. Moreover, it will be useful for various computational applications; in particular, the information we are encoding on cross-linguistic differences in lexicalization patterns will be crucial for applications like e.g. machine translation or language learning systems.

4. Concluding remarks

Aiming at a cost-effective multilingual resource, useful for various Language Engineering applications, within the EWN project we dealt with both theoretical and practical issues in order to define the set of relations to be encoded in the semantic database. The framework we developed is quite rich, although the information we are encoding is not, of course, *all* the

semantic information which could be encoded for lexical items. In any case, given the limitations (in time and budget) of the present project, we will not be able to express all the relations defined for the full vocabulary of each language in the EWN context. Indeed, the more 'sophisticated' relations, such as the Involved relation, are being encoded in cases in which they are easily extractable and/or really necessary to precisely locate a word meaning in the semantic network as a whole (especially when the traditionally accepted relations such as hyponymy and synonymy are not very helpful). We believe, however, that the database we are building could be further filled in and enriched in the future, also with data coming from other sources and, in particular, from textual corpora. This will be partly performed, for the Italian wordnet, within a national research project (starting in 1999), in which the results obtained in EWN will be extended by encoding further data in connection with both the same lexical subsets already taken into consideration within EWN and additional subsets. In any case, we believe that the framework we have designed can prove to be very useful for computational applications and could become a *model* for semantic resources of the same kind.

5. Notes

- ¹ The EWN project is a 3-year EC-funded project whose partners are currently the University of Amsterdam (coordinator); the Istituto di Linguistica Computazionale, CNR, Pisa; the Fundacion Universidad Empresa (a cooperation of UNED, Madrid, Politecnica de Catalunya, Barcelona, and the University of Barcelona); the University of Sheffield; and Novell Linguistic Development (Antwerp). In the future, the database will be extended with German, French, Estonian and Czech.
- ² Again as in Wordnet, we adopted a *weak* definition of synonymy, 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. 1993). That is, one such a context is sufficient to allow the identification of a synonymy relation.
- ³ The term *troponymy* is used in WN to indicate the relation occurring between a more general and a more specific concept when dealing with verbs, given that this relation has a somehow different nature with respect to the hyponymy relation occurring between nouns (Fellbaum 1993). However, since in both cases we are dealing with the logical notion of inclusion of one class in another, in EWN we use the same term *hyponymy* to identify the relation involving either nouns or verbs.
- ⁴ Here and in the following we use the term *verb* to refer to a specific *verb sense*. Furthermore, when discussing the relations encoded for *verbs* we actually refer to relations encoded for verb *synsets*, although we do not indicate whole synsets here.
- ⁵ Although some relations between a noun or a verb on one hand and an adjective or an adverb on the other are being encoded, we are not building complete wordnets for adjectives and adverbs.
- ⁶ The relation *Role* is used for the opposite link, from concrete nouns, or adjectives and adverbs, to verbs (or nouns referring to states, processes or events).
- ⁷ The notions of 'agent' and 'patient' we refer to are very similar to the notions of proto-agent/proto-patient defined by Dowty (1988). The notion of 'instrument' is used in a wide sense, to refer to any object used to do something: thus, also vehicles are encoded as 'instruments'. The notions of 'location' and 'direction' are being used assuming Gruber (1976) view (as further developed by Jackendoff 1983; 1990) according to which the semantics of motion and collocation can be seen as providing an interpretation for many semantic classes of verbs (besides the motion one), e.g. verbs of *bringing*, *saying*, *giving*.

- ⁸ Information on the *Acquilex* project can be found at the project's www homepage: <http://www.cl.cam.ac.uk/Research/NL/acquilex/acqhome.html>), while information on *Delis* can be found at: <http://www2.echo.lu/langeng/en/lre1/delis.html>.

6. References

- Alonge, A. (1992) Machine-Readable Dictionaries and Lexical Information on Verbs, in H. Tommola, K. Varantola, T. Salmi-Tolonen, J. Schopp (eds.), *Proceedings of the 5th Euralex International Congress*, Tampere, University of Tampere, pp. 195-202.
- Alonge, A. (1994) Motion Verbs: Data on Meaning Components in Dictionaries and Identification of Syntactic Properties, in W. Martin, W. Meijs, M. Moerland, E. ten Pas, P. van Sterkenburg, P. Vossen (eds.), *Proceedings of the 6th Euralex International Congress*, Amsterdam, pp. 27-35.
- Alonge, A. (1996) *Definition of the Links and Subsets for Verbs, EuroWordNet*, LE2-4003, Deliverable D006, Amsterdam, Computer Centrum Letteren, University of Amsterdam.
- Alonge, A., L. Bloksma, N. Calzolari, I. Castellon, A. Marti, W. Peters, and P. Vossen (forthcoming) The Linguistic Design of the EuroWordNet Database, *Computers and the Humanities: Special Issue on EuroWordNet*.
- Agirre, E. and G. Rigau (1996) Word Sense Disambiguation Using Conceptual Density, in *Proceedings of COLING*, Copenhagen, pp. 16-22.
- Bloksma, L., P. L. Diez-Orzas and P. Vossen (1996) *The User-Requirements and Functional Specification of the EuroWordNet Project, EuroWordNet*, LE2-4003, Deliverable D001, Amsterdam, Computer Centrum Letteren, University of Amsterdam.
- Climent, S., H Rodriguez and J. Gonzalo (1996) *Definition of the Links and Subsets for Nouns of the EuroWordNet project, EuroWordNet*, LE2-4003, Deliverable D005, Amsterdam, Computer Centrum Letteren, University of Amsterdam.
- Cruse, D. A. (1986) *Lexical Semantics*, Cambridge, Cambridge University Press.
- Dowty, D. R. (1988) Thematic Protoroles, Subject Selection, and Lexical Semantic Defaults, 1987 *LSA Colloquium Paper*.
- Fellbaum (1993) English Verbs as a Semantic Net, *Five Papers on WordNet*, Technical Report, Cognitive Science Laboratory, Princeton University.
- Gruber, J. (1976) *Lexical Structures in Syntax and Semantics*, Amsterdam, North-Holland.
- Jackendoff, R. (1983) *Semantics and Cognition*, Cambridge, Mass., The MIT Press.
- Jackendoff, R. (1990) *Semantic Structures*, Cambridge, Mass., The MIT Press.
- Levin, B. (1993) *English Verb Classes and Alternations*, Chicago, The University of Chicago Press.
- Levin, B. and M. Rappaport (1992) The Lexical Semantics of Verbs of Motion, in I. M. Roca (ed.), *Thematic Structure: its Role in Grammar*, Berlin - New York, Foris, pp. 246-267.
- Miller, G., R. Beckwith, C. Fellbaum, D Gross and K. Miller (1993) Introduction to WordNet: An On-Line Lexical Database, *Five Papers on WordNet*, Technical Report, Cognitive Science Laboratory, Princeton University.
- Richardson, R. and A. Smeaton (1995) Using WordNet in a Knowledge-Based Approach to Information Retrieval, in *Proceedings of the BCS-IRSG Colloquium*, Crewe.
- Sanfilippo, A. (1997) Using Semantic Similarity to Acquire Cooccurrence Restrictions from Corpora, in P. Vossen, G. Adriaens, N. Calzolari, A. Sanfilippo, Y. Wilks (eds.),

Proceedings of the ACL-EACL Workshop on 'Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications', Madrid, pp. 82-89.

Sanfilippo, A., T. Briscoe, A. Copestake, M. A. Martí Antonin, A. Alonge (1992) Translation Equivalence and Lexicalization in the ACQUILEX LKB, in *Proceedings of the 4th International Conference on Theoretical and Methodological Issues in Machine Translation*, Montreal, Canada.

Talmy, L. (1985) Lexicalization Patterns: Semantic Structure in Lexical Form, in T. Shopen (ed.), *Language Typology and Syntactic Description: Grammatical Categories and the Lexicon*, Cambridge, Cambridge University Press, pp. 57-149.