VU Research Portal

EuroWordNet: final report
Vossen, P.J.T.M.

1999

document version
Publisher's PDF, also known as Version of record

Link to publication in VU Research Portal

citation for published version (APA)
Vossen, P. J. T. M. (1999). *EuroWordNet: final report*. Vrije Universiteit.

EuroWordNet Final Report

Piek Vossen, University of Amsterdam

Version 2, 1999
Final

Deliverable D041, Work Package 0
EuroWordNet, LE2-4003, LE4-8328

Identification number	LE2-4003-LE4-8328-D-041
Type	Document
Title	Final Report
Status	Draft
Deliverable	D041
Work Package	WP0
Task	n.a.
Period covered	March 1996 – June 1999
Date	January 14, 2004
Version	2
Number of pages	34
Authors	Piek Vossen, University of Amsterdam
WP/Task responsible	AMS
Project contact point	Piek Vossen Computer Linguistiek University of Amsterdam Spuistraat 134 1012 VB Amsterdam The Netherlands tel. +31 20 525 4669 fax. +31 20 525 4429 e-mail: Piek.Vossen@hum.uva.nl
EC project officer	Ray Hudson
Status	Public
Actual distribution	Project Consortium, EC, WWW
Suplementary notes	n.a.
Key words	EuroWordNet, Multilingual database, wordnets, semantics and information retrieval.
Abstract	Final Report
Status of the abstract	Final
Received on	
Recipient's catalogue number	

Executive Summary

The goal of EuroWordNet was to build a multilingual lexical database with wordnets for several European languages, which are structured along the same lines as the Princeton WordNet. WordNet contains information about nouns, verbs, adjectives and adverbs in English and is organized around the notion of a *synset*. A synset is a set of words with the same part-of-speech that can be interchanged in a certain context. For example, {car; auto; automobile; machine; motorcar} form a synset because they can be used to refer to the same concept. A synset is often further described by a gloss: "4-wheeled; usually propelled by an internal combustion engine". Finally, synsets can be related to each other by semantic relations, such as hyponymy (between specific and more general concepts), meronymy (between parts and wholes), cause, etc. as is illustrated in Figure 1.
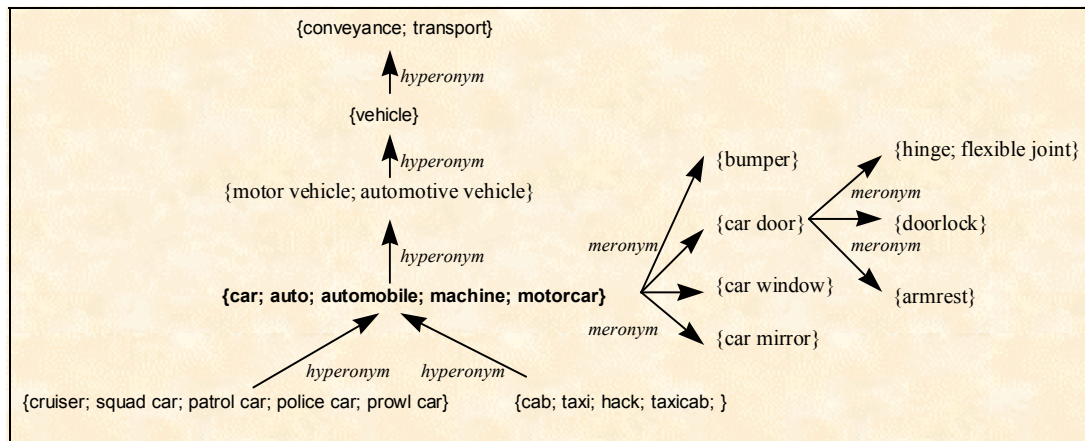

Figure 1: Synsets related to *car* in its first sense in WordNet1.

In this example, taken from WordNet1.5, the synset {car; auto; automobile; machine; motorcar} is related to:

- more general concepts or the hyperonym synset: {motor vehicle; automotive vehicle},
- more specific concepts or hyponym synsets: e.g. {cruiser; squad car; patrol car; police car; prowl car} and {cab; taxi; hack; taxicab},
- parts it is composed of: e.g. {bumper}; {car door}, {car mirror} and {car window}.

Each of these synsets is again related to other synsets as is illustrated for {motor vehicle; automotive vehicle} that is related to {vehicle}, and {car door} that is related to other parts: {hinge; flexible joint}, {armrest}, {doorlock}. By means of these and other semantic/conceptual relations, all meanings can be interconnected, constituting a huge network or wordnet. Such a wordnet can be used for making semantic inferences about the meanings of words (what meanings can be interpreted as *vehicles*), for finding alternative expressions or wordings, or for simply expanding words to sets of semantically related or close words in information retrieval. Furthermore, semantic networks give information on the lexicalization patterns of languages, on the conceptual density of areas of the vocabulary and on the distribution of semantic distinctions or relations over different areas of the vocabulary.

The European wordnets are stored in a central lexical database system and each meaning is linked to a so-called Inter-Lingual-Index, thus creating a multilingual database. This index is based on the concepts in WordNet1.5, but has been adapted to provide a more efficient mapping. In the multilingual database it is possible to go from one meaning in a wordnet to a meaning in another wordnet, which is linked to the same index-record. Such a multilingual database is useful for cross-language information retrieval, for transfer of information from one resource to another or for simply comparing the different wordnets. A comparison may tell us something about the consistency of the relations across wordnets, where differences may point

to inconsistencies or to language-specific properties of the resources, or also to properties of the language itself.

In EuroWordNet, we initially worked on 4 languages: Dutch, Italian, Spanish and English. The size of each of these wordnets, except for English, is about 30,000 synsets (roughly corresponding to 50,000 word meanings) with 2,05 up to 2,90 semantic relations on average between concepts. For comparison, the size of WordNet1.5 is 91,591 synsets and 168,217 word meanings. In an extension to the project, the database has been extended with German, French, Estonian and Czech: the size of these wordnets is between 7,000-20,000 synsets.

Through the Inter-Lingual-Index, the wordnets share a top-ontology that has been applied to the most fundamental concepts: the so-called Base Concepts. These 1300 Base Concepts play an important role in establishing the semantic relations in wordnets for different languages. The wordnets have been built according to a common top-down approach that ensured maximum compatibility and flexibility. The same set of Base Concepts has been used to first develop core wordnets that are highly compatible and of high quality, with a rich density of relations. These core wordnets have been extended to cover more specific concepts. All the wordnets are provided as plain text files and in a database format. The database versions can be accessed, edited and compared in the multilingual database Polaris or viewed with the graphical interface Periscope. The wordnets are distributed via ELDA/ELRA.

In addition to the builders, there have been 3 industrial users in the project, where Novell also had an additional role as the developer of the shared EuroWordNet database.:

• Bertin & Cie, Plaisir, France
• Xerox Research Centre, Meylan, France
• Novell Linguistic Development, Antwerp, Belgium (replaced by Lernout & Hauspie in the final year).

They verified the quality and coverage of the data and demonstrated the use of the database in their (multilingual) information-retrieval applications. The experiments with the French and German wordnet showed that substantial improvements are made in a multilingual retrieval context (English, German and French), this despite the small size of the wordnets and the lack of word sense disambiguation (simply all senses have been considered). Monolingual retrieval has only been applied to French. Here no improvements have been reported compared to baseline retrieval. It is not clear to what extent this relates to the small size of the French wordnet and the lack of word sense disambiguation.

Further validation and feedback has been done internally, by developing specific comparison options to measure compatibility of wordnets. External feedback was done by 3 project-reviews and by direct dissemination of the results to a user-group of 70 institutes and companies that have expressed their interest in the project.

The wordnets represent basic resources for content-based language-technologies within and across the languages. The generic wordnets can be extended semi-automatically for specific domains. As a multilingual database, EuroWordNet can be used to share these technologies across any associated language. In addition to the use for (cross-language) information retrieval, there are many other applications that can directly benefit from the multilingual semantic resources: information-acquisition tools, authoring-tools, language-learning tools, translation-tools, summarizers.

Finally, EuroWordNet represents a framework for standardizing lexical semantic resources in languages, and eventually, for developing a standardized index of meaning that can be used for developing and comparing content-based applications in any language that is linked to the index. There are currently many institutes developing wordnets according to the EuroWordNet specification. Wordnets developed in collaboration with EuroWordNet cover the following languages: Basque, Catalan, Portuguese, Danish, Lithuan, Swedish, Russian, Greek. In this respect, we are investigating the possibility to start a Global Wordnet Association to maintain the framework.

Table of Contents

1. Introduction

This document summarizes the results and prospects of EuroWordNet, which was funded under two contracts: LE2-4003 and LE4-8328. The aim of EuroWordNet was the development of a multilingual database with semantic networks for 8 European languages: English, Dutch, Spanish, Italian, German, French, Czech and Estonian. The semantic networks are structured along the same lines as the American WordNet developed at Princeton University. Synonymous words are organized as so-called synsets, and semantic relations such as hyponymy, meronymy, cause and role, are expressed between these synsets. Synsets in the individual languages are interconnected via a so-called Inter-Lingual-Index. The resulting multilingual database was tested in monolingual and cross-lingual information retrieval applications.

The first contract LE2-4003, referred to as EuroWordNet-1 or EWN1, covered the languages English, Dutch, Spanish and Italian and the development of the EuroWordNet database. After a successful start the project was extended with a second contract LE4-8328, referred to as EuroWordNet-2 or EWN2, to include wordnets for the languages French, German, Czech and Estonian.

Section 2 of this report covers the contractual aspects, the stages of work and major events for both contracts. Section 3 describes the major achievements, and section 4 the evaluation and assessment. Finally, conclusions, future aspects and the exploitation plans are discussed in section 5. The appendices list the deliverables, publications and the user-group contacts.

2. Contractual Aspects

EWN1 and EWN2 have been funded by the European Commission, DG XIII, Luxembourg as projects LE2-4003 and LE4-8328 in the application areas: Language Resources and Language Engineering.

LE2-4003

Title:	EuroWordNet: Building a multilingual wordnet database with semantic relations between words
Start Date:	1-March-1996
End Date:	30-June-1999
Duration:	36 months
Global Effort:	149 person months

LE4-8328

Title:	EuroWordNet-2: Extending EuroWordNet with other languages.
Start Date:	9-April-1998
End Date:	8-July-1999
Duration:	15 months
Global Effort:	137,5 person months

Organization	*Nat.*	*Task*	*Contract*
University of Amsterdam	NL	Coordinator&Dutch wordnet	LE2-4003&LE4-8328
Istituto di Linguistica Computazionale, CNR, Pisa	IT	Italian wordnet	LE2-4003&LE4-8328
Fundacion Universidad Empresa	ES	Spanish wordnet	LE2-4003&LE4-8328
Université d' Avignon and Memodata at Avignon	FR	French wordnet	LE4-8328
Universität Tübingen	DE	German wordnet	LE4-8328
University of Masaryk at Brno	CZ	Czech wordnet	LE4-8328
University of Tartu, Estonia	EE	Estonian wordnet	LE4-8328
University of Sheffield	GB	Adapt the English wordnet	LE2-4003&LE4-8328
Novell Belgium NV	BE	User & database developm.	LE2-4003
Lernout and Hauspie	BE	User database developm.	LE2-4003
Xerox Research Centre, Meylan	FR	User	LE4-8328
Bertin & Cie, Plaisir, Paris	FR	User	LE4-8328

	Time Schedule EWN1, LE2-4003	1996				1997				1998			1999
		Mar	June	Sep	Dec	Mar	June	Sep	Dec	Mar	June	Sep	Dec
		m1-3	m4-6	m7-9	m10-12	m13-15	m16-18	m19-21	m22-24	m25-27	m28-30	m31-33	m34-36
WP0	Management	■	■	■	■	■	■	■	■	■	■	■	■
WP1	User requirements and functional specification	■											
WP2	Tools and resources	■	■	■	■			■	■				
WP3	Build Noun Wordnets		■	■	■	■	■		■	■			
WP4	Build Verb Wordnets		■	■	■	■	■		■	■			
WP5	Top-ontology						■	■	■	■	■	■	
WP6	EuroWordNet Database		■	■	■	■	■		■				
WP7	Validation					■	■	■	■	■			■
WP8	Exploitation plan								■				■
WP9	Awareness and dissemination*	■	■	■				■					■
WP10	Concertation*	■						■		■			
	Reviews					R			R				R

	Time Schedule, EWN2, LE4-8328	1998									1999					
		Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec	Jan	Feb	Mar	Apr	May	Jun
		m1	m2	m3	m4	m5	m6	m7	m8	m9	m10	m11	m12	m13	m14	m15
Wp0	Management	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
Wp1	Functional Specification	R	R	R												
Wp2	Tools and Resources	R	R	R	R	R	R									
Wp3	Build Noun Wordnets	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
Wp4	Build Verb Wordnets	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
Wp7	Validation										R	R	R	R	R	R
WP10	Concertation	R	R	R	R	R	R	R	R	R	R	R	R	R	R	R
WP11	Adapt the Inter-Lingual-Index	R	R	R	R	R	R	R	R	R	R	R				
	Reviews											■				

Specification phase:
The project EWN-1 (LE2-4003) started with a specification phase, which covered:

- specification of the user-requirements;
- the design of the database;
- specification of the semantic relations with verification tests;
- conduction of experiments to verify the design;
- the specification of first selections of the vocabulary;
- a common methodology for the top-down development of the wordnets;
- preparation and adaptation of the local databases, resources and tools;

The first phase has been completed in time at T7. The main deliverables are:

> *D001: User requirements and functional specification*
> *D005: Definition of the links and subsets for nouns*
> *D006: Definition of the links and subsets for verbs*
> *D007: Architecture of the EuroWordNet database*

The specification and design of the EuroWordNet database was presented at the first project review in February 1997 and at the ACL/EACL post-conference workshop, Madrid-1997. This workshop was organized by EuroWordNet in collaboration with two other LE1 projects Sparkle and Ecran.

Building phase-1, EWN1:
Using local tools and databases we developed the core wordnets around a set of so-called Base Concepts, which are the most important concepts in the wordnets. These Base Concepts have been classified by a language-neutral top-ontology, developed specifically for this purpose. This ensured a maximum of compatibility and overlap of the core wordnets.

A side effect of this approach has been that we had to adapt the work plan. Originally, the development of the top-ontology was foreseen at the end of each building phase. Now this work was moved forward to provide a common framework, consequently, delaying the building of the first wordnets. Instead of June 1997, the first subset was finished in December 1997. Also the development of the EuroWordNet database was delayed. The department of Novell in Antwerp was taken over by Lernout and Hauspie, but during the negotiation phase, which lasted 1,5 year, no extra investments could be made. The validation task by Novell was dropped and all funds have been used to develop the EuroWordNet database Polaris. The first version was released in September 1997 and the core wordnets could be compared and restructured in the beginning of 1998, just before the second review (February 1998). At that time Lernout and Hauspie took over the role of Novell to support the EuroWordNet database. To compensate the verification task, we carried out an in-depth comparison of wordnets fragments in the database, and developed a system for comparing the overall structures and monitoring the progress of the wordnets in other databases.

The major deliverables are:
> *D008: Multilingual storage and viewing*
> *D010D011: Subset1 for Dutch, Spanish, Italian and English*
> *D014D015: Restructured subset1 for Dutch, Spanish, and Italian*
> *D017D034D036: The EuroWordNet Top-Ontology*
> *D024: EuroWordNet Database Report*
> *D025: EuroWordNet Tools Report*

Building phase-2, EWN1:
For the second building phase, we used the results of the comparison to direct the extension and the improvements. Furthermore, the wordnets have been compared with the Parole lexicons in the same languages and with corpus frequency. The second building phase consisted of separate extensions and two more comparisons of the data. The focus has been on improvement of the equivalence relations, on adding

non-hyponymy relations and increasing the overlap across the wordnets. The final wordnets have been completed in the first half-year of 1999.

The major EWN1 deliverables are:
> *D027D028: Subset2 for Dutch, Spanish, Italian and English*
> *D029D030: Restructured subset2 for Dutch, Spanish, and Italian*
> *D032D033: The Wordnet Report for Dutch, Spanish, Italian and English*

EWN2 started after the second review of EWN1.

Specification phase EWN2:
The specification of the EuroWordNet database was verified by the builders of the French, German, Czech and Estonian wordnets. For this task the new partners encoded small samples of concepts manually. The verification was very positive. No changes were needed in the design, except for a minor adaptation of the database to deal with other character sets.

Another verification task was carried out by specifying independent sets of Base Concepts in the new languages, using similar criteria as in EWN1. These new selections have been compared within EWN2 and also with the set of common Base Concepts derived in EWN1. This showed that most of the common Base Concepts (up to 85%) were covered in both selections. Some new concepts have been added to the set of common Base Concepts. The extended set of common Base Concepts has then been used to first develop the core wordnets.

Building phase-1 EWN2:
Next, the new partners followed the same approach as in EWN1. They first developed core wordnets around the Base Concepts. The core wordnets have been loaded in the EuroWordNet database and have been compared in the same way as has been done in EuroWordNet-1. Comparison has been done by the new EWN2 partners and by EWN1 partners.

The work for EWN1 showed that the Inter-Lingual-Index (ILI), which was initially based on the Princeton WordNet1.5, had to be adapted to provide a more efficient mapping between wordnets. In EWN2, a set of heuristics was developed that clustered closely related senses. These clustered index items have been added to the ILI as so-called Composite ILI-records (4,608 in total, grouping 8,339 concepts)). The links from the wordnets were automatically updated for these new records. This increased the overlap across wordnets with 5% for nouns and 100% for verbs.

The major EWN2 deliverables for this phase are:
> *2D001: The revised set of common Base Concepts*
> *2D002: Specification of German & French WNs*
> *2D003: Specification of Czech & Estonian WNs*
> *2D004: The restructured Inter-Lingual-Index*
> *2D005: Tools & resources German & French WNs*
> *2D006: Tools & resources Estonian & Czech WNs*
> *2D007: First WNs for BCs for French, German, Czech & Estonian*
> *2D008: Compared Subset1 for French, German, Czech & Estonian*

Building phase-2 EWN2:
For the first subset of EWN2, an independent verification of the core wordnets for French and German was carried out by the user in the project. The input of the verification and the comparison has then been used to complete the wordnets, which have been delivered in time before the summer of 1999. Furthermore, we have studied the possibility to extend and further restructure the ILI on the basis of synsets in non-English languages that could not be mapped to the ILI. This discussion has resulted in the development of a model for a universal index of meaning, which was presented at the ACL-99 SIGLEX workshop in Maryland. The development of such a minimized and universal ILI requires some adaptations to the EuroWordNet

database that could not be implemented in the remaining time frame and budget of the project. Such a task has to be carried out in a new project.

The deliverables for this period are:
> *2D009: Verification of the core wordnets for French & German*
> *2D010: Extended Inter-Lingual-Index*
> *2D011D012: Comparison of the final wordnets for German, French, Czech and Estonian*
> *2D014: Wordnet document for French, German, Estonian and Czech*

3. Achievements

The EuroWordNet-database is first of all based on the structure of the Princeton WordNet and specifically version WordNet1.5. The notion of a synset and the main semantic relations have been taken over in EuroWordNet. However, some specific changes have been made to the design of the database, which are mainly motivated by the following objectives:

1) to create a multilingual database;
2) to maintain language-specific relations in the wordnets;
3) to achieve maximal compatibility across the different resources;
4) to build the wordnets relatively independently (re)-using existing resources;

In the design of the database, we make a distinction between the language-specific modules and a separate language-independent module. Each language module represents an autonomous and unique language-specific system of language-internal relations between synsets. Equivalence relations between the synsets in different languages and WordNet1.5 will be made explicit in the so-called Inter-Lingual-Index (ILI). Each synset in the monolingual wordnets will have at least one equivalence relation with a record in this ILI. Language-specific synsets linked to the same ILI-record should thus be equivalent across the languages, as is illustrated in Figure 2 for the language-specific synsets linked to the ILI-record *drive*.

Figure 2 further gives a schematic presentation of the different modules and their inter-relations. In the middle, the language-external modules are given: the ILI, a Domain Ontology and a Top Concept Ontology. The language-internal modules then consist of a lexical-item-table indexed to a set of word-meanings, between which the language-internal relations are expressed. The ILI is an unstructured list of meanings, mainly taken from WordNet1.5, where each ILI-record consists of a synset, an English gloss specifying the meaning and a reference to its source. The only purpose of the ILI is to mediate between the synsets of the language-specific wordnets. No relations are therefore maintained between the ILI-records as such. Hierarchical structuring of the ILI-records is given by each wordnet that is linked to it, including WordNet1.5. Some language-independent structuring of the ILI is nevertheless provided by two separate ontologies, which are linked to ILI records:

- the Top Concept ontology, which is a hierarchy of language-independent concepts, reflecting important semantic distinctions, e.g. Object and Substance, Location, Dynamic and Static;
- a hierarchy of domain labels, which are knowledge structures grouping meanings in terms of topics or scripts, e.g. Traffic, Road-Traffic, Air-Traffic, Sports, Hospital, Restaurant;

Both the Top Concepts and the domain labels can be transferred via the equivalence relations of the ILI-records to the language-specific meanings, as is illustrated in Figure 2. The Top Concepts *Location* and *Dynamic* are for example directly linked to the ILI-record *drive* and therefore indirectly also apply to all language-specific concepts related to this ILI-record. Via the language-internal relations the Top Concept can be further inherited by all other related language-specific concepts. The main purpose of the Top Ontology is to provide a common framework for the most important concepts in all the wordnets. It consists of 63 basic semantic distinctions that classify a set of 1300 ILI-records representing the most important concepts in the different wordnets. The classification has been verified by the different sites, so that it holds for all the language-specific wordnets. The domain-labels can be used directly in information retrieval (and also in language-learning tools and dictionary publishing) to group concepts in a different

way, based on scripts rather than classification. Domains can also be used to separate the generic from the domain-specific vocabularies. This is important to control the ambiguity problem in Natural Language Processing.
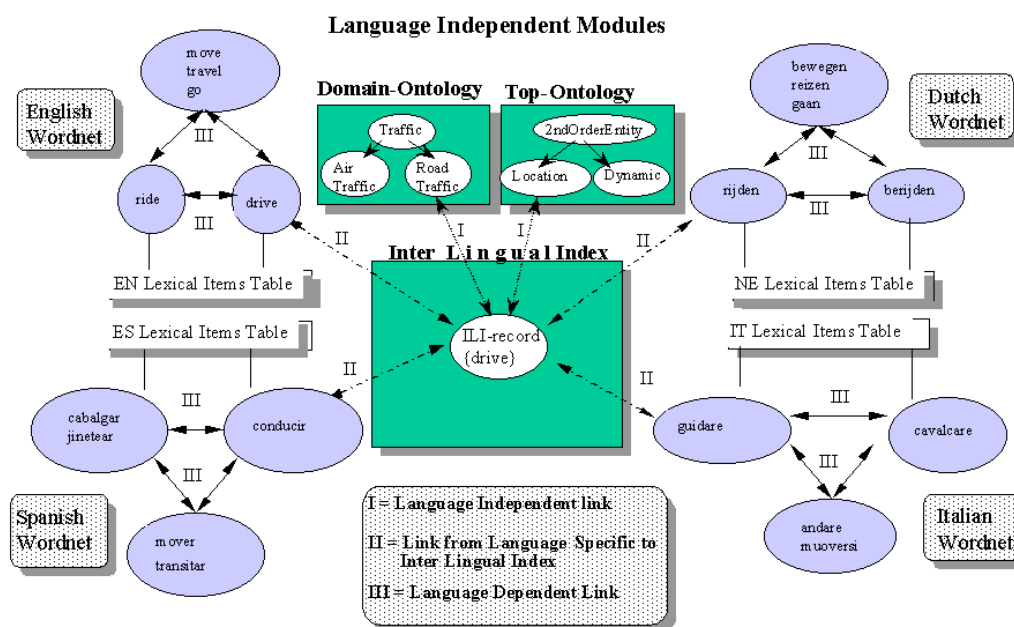

Figure 2. The global architecture of the EuroWordNet database.

The EuroWordNet database makes it possible to compare wordnet fragments via the ILI and to track down differences in lexicalization and in the language-internal relations. This is illustrated in Figure 3, which is taken from the graphical interface to the EuroWordNet database, called Periscope. The top-half of the screen-dump shows a window with a fragment of the Dutch wordnet at the left and a similar fragment of WordNet1.5 at the right. The bottom window shows a similar parallel view for the Italian and Spanish wordnets. Each synset in these windows is represented by a rectangular box followed by the synset members. On the next line, the closest Inter-Lingual-Index concept is given, following the = sign (which indicates direct equivalence). In this view, the ILI-records are represented by an English gloss. Below a synset-ILI pair, the language-internal relations can be expanded, as is done here for the hyperonyms. The target of each relation is again represented as a synset with the nearest ILI-equivalent (if present). The first line of each wordnet gives the equivalent of *cello* in the 4 wordnets. In this case, they are all linked to the same ILI-record, which indirectly suggests that they should be equivalent across the wordnets as well. We also see that the hyperonyms of *cello* are also equivalent in the two windows, as is indicated by the lines connecting the ILI-records. Apparently, the structures are parallel across the Dutch wordnet and WordNet1.5 on the one hand and the Spanish and Italian wordnets on the other. However, we see that the intermediate levels for *bowed stringed instrument* and *stringed instrument* in the Dutch wordnet and WordNet1.5 are missing both in Italian and Spanish. Had we compared other wordnet pairs, the intermediate synsets would be unmatched across the wordnets.
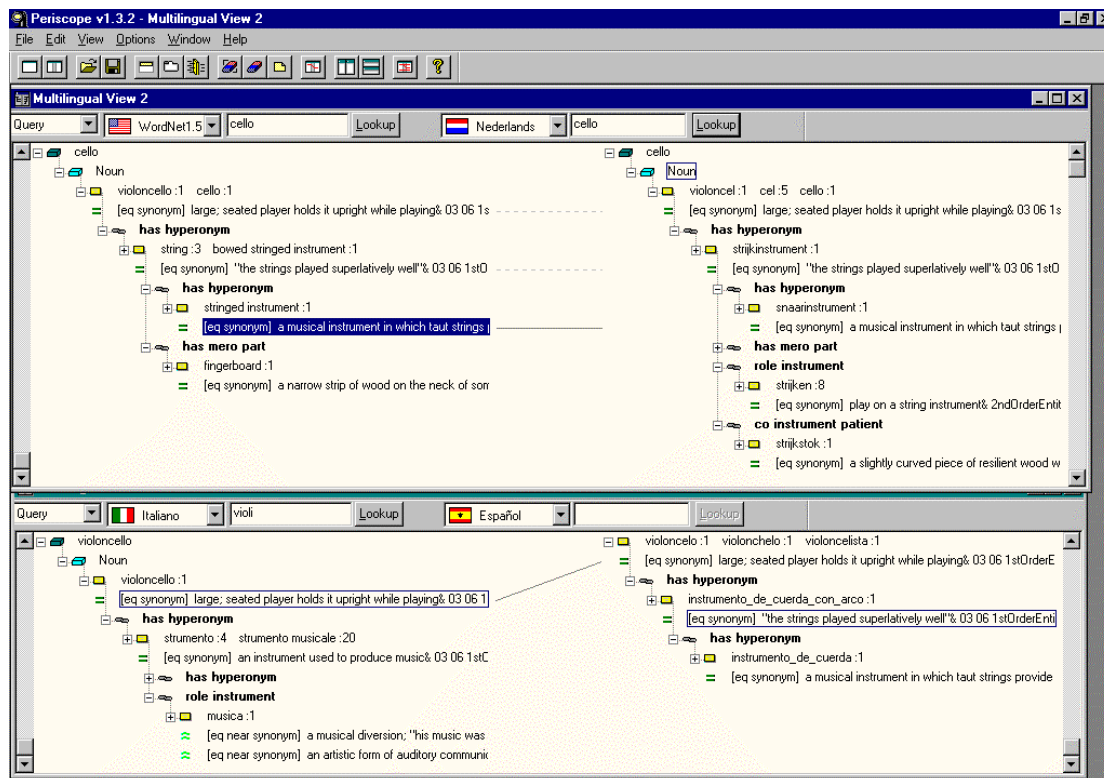
Figure 3: Parallel wordnet structures in EuroWordNet linked to the same ILI-records.

The modular multilingual design of the EWN-database has the following advantages:

- it is possible to use the database for multilingual information retrieval, by expanding words in one language to related words in another language via the ILI;
- the different wordnets can be compared and checked cross-linguistically;
- language-dependent differences are maintained in the individual wordnets;
- it is possible to develop new wordnets or extend existing ones relatively independently;
- language-independent information such as the glosses, the domain-knowledge and the analytic Top Concepts are stored only once and are made available to all the language-specific modules via the inter-lingual relations;
- the database can be tailored to a user's needs by modifying the Top Concepts, the domain labels or instances, (e.g. by adding semantic features) without having to access the language-specific wordnets;

In addition to the multilingual design of the database, there have been some changes to the language-internal relations with respect to WordNet1.5. The major innovations are:

1. the use of labels to relations, which makes the semantic entailments more explicit and precise;
2. the introduction of cross part-of-speech relations, so that different surface realizations of similar concepts within and across languages can still be matched;
3. the addition of additional relations to differentiate certain shallow hierarchies;

Conjunction and *disjunction* are examples of relation labels that can be assigned to multiple relations of the same kind. Conjunction of relations is typical for meronymy: i.e. multiple parts that together make up a single whole (e.g. *wings*, *nose*, *tail*, *door* are parts that make up an *airplane* conjunctively). However, it may also apply to other relations such as hyponymy: a *knife* is both a *weapon* and a *piece of cutlery* at the same time. In other cases, multiple parts or hyperonyms are clearly disjunctive: an *albino* is either an *animal*, *human* or a *plant*, a *threat* may be a *person*, *idea* or *thing*, an *airplane* either has *propellers* or *jets*.

Whereas in WordNet, the parts-of-speech represent distinct networks, in EuroWordNet they are interconnected in various ways. A typical cross-part-speech relations is xpos-synonymy between words with different part-of-speech that can be used to describe the same concept, e.g. between the verb *adorn* and the noun *adornment* or the noun *death* and the adjective *dead*. However, also other relations across parts-of-speech are allowed such as: causation relations between *die* and *dead*, *redden* and *red*, or semantic role relations between nouns and verbs, such as agent (*teacher*), patient (*student*), location (*school*) related to *teach*. The latter relations also differentiate shallow hierarchies, such as *persons* or *physical changes*, where many hyponyms only differ in the associated role or result.

These changes directly improve the use of the database for Language Engineering applications. In total 90 different language-internal relations have been specified. All the language-internal relations have been defined using explicit tests in all the four languages. These tests ensure minimal consensus on the interpretation of the relations across the sites. In addition, 20 types of equivalence relations have been distinguished. The next table gives a quantitative overview of the final wordnets.

Explanation of the columns:

Synsets	= concepts represented by synonymous word senses
No. of senses	= number of word senses, or synonyms
Sens./ syns.	= average of senses or synonyms per synset
Entries	= number of words
Sens./ entry	= number of senses per word
LIRels.	= number language-internal relations
LIRels/ syns	= average of language-internal relations per synset
EQRels-ILI	= number of equivalence relations
EQRels/syn	= average of equivalence relations per synset
Synsets without ILI	= synsets without a equivalence relation
%without ILI	= percentage of synsets without an equivalence relations

Quantitative overview of the EuroWordNet database

		Synsets	No. of senses	Sens./ syns.	Entries	Sens./ entry	LIRels.	LIRels/ syns	EQRels-ILI	EQRels /syn	Synsets without ILI	%without ILI
Dutch Wordnet	Nouns	34455	54428	1,58	45972	1,18	84869	2,46	26724	0,78	6070	17,62%
	Verbs	9040	14151	1,57	8826	1,60	25973	2,87	26724	2,96	1133	12,53%
	Other	520	1622	3,12	1485	1,09	797	1,53	n.a.	n.a.	n.a.	n.a.
	Total	44015	70201	1,59	56283	1,25	111639	2,54	53448	1,21	7203	16,36%
Spanish Wordnet	Nouns	18577	41292	2,22	23216	1,78	40559	2,18	18634	1,00	0	0,00%
	Verbs	2602	6795	2,61	2278	2,98	3749	1,44	2602	1,00	0	0,00%
	Other	2191	2439	1,11	2439	1,00	10855	4,95	n.a.	n.a.	n.a.	n.a.
	Total	23370	50526	2,16	27933	1,81	55163	2,36	21236	0,91	0	0,00%
Italian Wordnet	Nouns	30169	34552	1,15	24903	1,39	83021	2,75	43848	1,45	98	0,32%
	Verbs	8796	12473	1,42	6607	1,89	30757	3,50	27941	3,18	0	0,00%
	Other	1463	1474	1,01	1468	1,00	3290	2,25	n.a.	n.a.	n.a.	n.a.
	Total	40428	48499	1,20	32978	1,47	117068	2,90	71789	1,78	1561	3,86%
French Wordnet	Nouns	17826	24499	1.37	14879	1.65	39172	2.20	17815	1.00	16	0.09%
	Verbs	4919	8310	1.69	3898	2.13	10322	2.10	4915	1.00	4	0.08%
	Other	0	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Total	22745	32809	1.44	18777	1.75	49494	2.18	22730	1.00	20	0.09%
German Wordnet	Nouns	9951	13656	1.37	12746	1.07	23856	2.40	10570	1.06	0	0.00%
	Verbs	5166	6778	1.31	4333	1.56	10960	2.12	5762	1.12	0	0.00%
	Other	15	19	1.27	19	1.00	2	0.13	15	1.00	0	0.00%
	Total	15132	20453	1.35	17098	1.20	34818	2.30	16347	1.08	0	0.00%
Czech Wordnet	Nouns	9727	13829	1.42	9277	1.49	19856	2.04	9729	1.00	0	0.00%
	Verbs	3097	6120	1.98	3006	2.04	6403	2.07	3097	1.00	0	0.00%
	Other	0	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Total	12824	19949	1.56	12283	1.62	26259	2.05	12824	1.00	0	0.00%
Estonian Wordnet	Nouns	5028	8226	1.64	7209	1.14	10873	2.16	5683	1.13	0	0.00%
	Verbs	2650	5613	2.12	3752	1.50	5445	2.05	3321	1.25	0	0.00%
	Other	0	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.	n.a.
	Total	7678	13839	1.80	10961	1.26	16318	2.13	9004	1.17	0	0.00%
English WordNet Addition	Nouns	4751	14188	2,99	2524	5,62	20707	4,36	n.a.	n.a.	n.a.	n.a.
	Verbs	11363	25761	2,27	14726	1,75	21070	1,85	n.a.	n.a.	n.a.	n.a.
	Other	247	639	2,59	70	9,13	363	1,47	n.a.	n.a.	n.a.	n.a.
	Total	16361	40588	2,48	17320	2,34	42140	2,58	n.a.	n.a.	n.a.	n.a.
WordNet1.5	Nouns	60521	107428	1,78	88175	1,22	159223	2,63	n.a.	n.a.	n.a.	n.a.
	Verbs	11363	25768	2,27	14734	1,75	24331	2,14	n.a.	n.a.	n.a.	n.a.
	Other	22631	54406	2,40	23708	2,29	27821	1,23	n.a.	n.a.	n.a.	n.a.
	Total	94515	187602	1,98	126617	1,48	211375	2,24	n.a.	n.a.	n.a.	n.a.

The wordnets are delivered as plain text files (according to the EuroWordNet format) and as EuroWordNet database files. The database files can be viewed with the public viewer Periscope or with the wordnet editor Polaris. As a database, the wordnets can be accessed cross-linguistically, structures can be compared, semantic selections can be projected from one language to another, data can be imported, exported and edited. In the Appendix, some screen-dumps are given from the Polaris database that illustrate how the wordnets can be accessed.

Both Periscope and Polaris run on Windows95/98/NT machines. The wordnets require between 10 and 25 MB disk space each. Another 70MB is needed for WordNet1.5 and the Inter-Lingual-Index. All data can however also be accessed from CD.

The wordnets represent basic resources for content-based language-technologies within and across the languages. The generic wordnets can be extended semi-automatically for specific domains. As a multilingual database, EuroWordNet can be used to share these technologies across any associated language. In addition to the use for (cross-language) information retrieval, there are many other applications that can directly benefit from the multilingual semantic resources: information-acquisition tools, authoring-tools, language-learning tools, translation-tools, summarizers. Furthermore, they can be used for enabling technologies such as: word-sense-disambiguation, improve speech recognition, spelling checkers, parsers, language-generation tools.

A final aspect to be mentioned is that EuroWordNet can be seen as a standardization project as well. The definition of the relations, the common data structure, the shared ontology, the Inter-Lingual-Index and the comparison option, have not only led to uniform structures across the 8 EuroWordNet languages, but are also taken over by many other sites outside the project. Currently, wordnets are developed for many other languages according to the EuroWordNet specification. The results of EuroWordNet have also been used and integrated in EAGLES, Simple and the ANSI committee for standardized ontologies.

4. Evaluation and assessment

Validation in EuroWordNet was organized in 3 phases:

- specification of the user-requirements and test-specifications;
- verification of the results of the first building phase
- integration and demonstration of the final results in local applications

The first phase was carried out by Novell Belgium (Antwerp), resulting in:

> *D001: User requirements and functional specification*
> *D007: Architecture of the EuroWordNet database*
> *D013: Test specification for EuroWordNet*

Furthermore, the database itself imposed various structural constraints on the data. Loading the data in the database requires a certain level of quality and uniformity. The actual verification of the EWN1 data has not been carried out by Novell, due to reasons explained above. As an alternative we have developed a method to do a further internal evaluation of the data by the builders, consisting of 3 modules:

- in-depth comparison of specific semantic clusters in the EuroWordNet database;
- overall comparison of the wordnets via top-ontology clustering;
- overall comparison of the hyponymy relations imposed on the ILI-records;

Within the database, it is possible to project semantic clusters to another wordnet and compare these with equivalent clusters. For example, *vehicles* in Spanish can be projected to the Dutch wordnet resulting in a list of Dutch synsets linked to the same ILI-records. These Dutch synsets can be compared with the *vehicles* in Dutch, e.g. extracted from the hyponymy relation in the Dutch wordnet. In the ideal case, these sets should be the same. Differences can be due to errors in equivalence relations and language-internal relations or to differences in coverage.

The second comparison is done in a local database of the University of Amsterdam. All the ILI-records associated with the synsets in a language can be imported in this database and the top-ontology concepts that apply to these records according to the WordNet1.5 hierarchy can be extracted. This results in a clustering of the associated ILI-records in terms of top-ontology concepts. Such a clustering is useful to see what areas of the top-ontology have been covered by a local wordnet, how well they are balanced and if there are differences in global coverage across the wordnets.

The third comparison is carried out by the University Politecnica de Catalunya in Barcelona, who developed a graph-mapping tool that compares the coverage and relations of so-called ILI-chains. These ILI-chains are derived from the local wordnets by imposing the hyponymy relation on the ILI-records that are associated with the synsets. This comparison gives a rough indication of the compatibility of the hyponymy relations across all the wordnets.

In the case of all 3 methods of comparison, the result does not only depend on the language-internal relations of the wordnets but also on the quality and coverage of the equivalence relations. These three comparison options have been applied to all the wordnets at different phases (some comparisons have been applied several times). The results have been published in the deliverables of each subset fragment and we

have based the continuation of the building on the results. As mentioned above, the vocabularies in the wordnets have also been checked in terms of overlap with the most frequent entries in the Parole lexicons and the most frequent words in corpora. This has also lead to extensions of the wordnets to guarantee the inclusion of the most frequent words.

For French and German, Xerox and Bertin have carried out an independent verification of the core wordnets, delivered after the first building phase. Internal quality tests have been performed as specified in the D013 deliverable. Internal testing of EuroWordNet consists of several generic checks that evaluate the integrity and consistency of the data, covering statistics, syntax and content. The verification extracted some errors, but none were very harmful, and all could easily be corrected.

After completion, the French and German wordnet have been integrated in the (cross-lingual) information retrieval tools of two users for validation. The associated deliverable is:

2D015: Demonstration of German and French wordnet in multilingual IR task

Xerox and Bertin tested the use of the database in their (multilingual) information-retrieval applications. The experiments done by Xerox with the French and German wordnet showed that substantial improvements are made in a multilingual retrieval context (English, German and French), this despite the small size of the wordnets and the lack of word sense disambiguation (simply all senses have been considered). Improvements are reported up to 20% and 24%, depending on the size of the queries, the combination of languages, and whether the EuroWordNet resources are combined with bilingual dictionaries. The latter is needed because of the limited size of the wordnets. Monolingual retrieval has only been done for French by Bertin. Here no improvements have been reported compared to baseline retrieval. It is not clear to what extent this relates to the small size of the French wordnet and the lack of word sense disambiguation.

Separately from the project-internal validation and verification, we have published the results to the user-group members, in various papers and publications and at workshops and conferences. The user-group has grown to 70 members (45 Universities and Research Institutes, 25 Companies and end-users), and many requests to use the data have been received by the separate partners. Globally, we can distinguish the following parties:

- publishers, either providing the initial resources or interested in the development of similar products;
- research institutes and R&D departments of universities and companies working in the field of knowledge engineering or linguistic databases, interested in building similar resources
- research institutes and R&D departments of universities and companies working in the field of Language Engineering interested in using or applying similar resources or developing services or products that make use of multilingual semantic resources: especially information processing;
- end-users interested in products helping them to deal with information;

The complete list with users that have returned a user-application form is given in the Appendices below. All the users have received the most important public deliverables and a free CD with samples of the wordnets and the graphical viewer Periscope (also distributed for free at the LREC conference in Granada in 1998), a specification of the Base Concepts and the top-ontology. All this information is also available on the EuroWordNet WWW-site, http://www.hum.uva.nl/~ewn.

Another major dissemination activity has been the publication of a special issue in the Computer and the Humanities journal in 1998 and a reprint of the special issue as a book by Kluwer. Furthermore, we have presented papers at all major conferences and platforms in Europe and the US, among which: ACL, EACL, LREC, Coling, Euralex, AAAI, IJCAI, Delos, Trec, TIA (see publication list below).

The major concertation activities have been:

- contribution to the EAGLES working group on lexical semantics;
- collaboration with other EC projects: Parole, Simple, Trevis;
- the organization of a joint ACL workshop with Sparkle and Ecran in 1997, Madrid;
- collaboration with ELRA, who acts as the distributor of the EuroWordNet results;
- participation in the ad hoc ANSI committee on ontology standards;

In addition, to these dissemination and concertation activities, we had 3 yearly reviews by external experts. The specification and design of the database has been presented at the first review after one year. The results of the building phases of EWN1 and EWN2 have been presented at the second and final project review. At the final review, Lernout & Hauspie also gave a demonstration of their cross-lingual information retrieval tool that exploits the multilingual wordnet database.

In all these occasions we received extremely positive feedback on the project. In so far there were comments, it is that we did not receive sufficient means and time to complete the job. Many experts, colleagues and potential users regret that we did not have the time and resources for this. Requested extensions are:

-    include adjectives and adverbs;
-    further improve the equivalence mapping,
-    further develop the ILI,
-    include multi-words and expressions,
-    integrate EuroWordNet and Parole,

Internally, we think that one of the bottlenecks is the quality of the equivalence mappings. Here further work and evaluation is needed. There are also many possibilities to build semi-automatic techniques using the combinations of wordnets and bilingual resources (e.g. Spanish-Italian).

5. Future prospects and exploitation

We believe that EuroWordNet is a solid fundament for the development of language resources and technology that can be shared and transferred to all the associated languages. The core wordnets have been developed carefully. Even though the size of the wordnets and the coverage of the non-hyponymy relations may not be complete for direct commercial application, these core wordnets will make it easy to extend and complete the data, possibly with semi-automatic techniques. The wordnets contain many features and structures that are unique in the world. These will provide many possibilities for experimentation in language technology. Examples are the relation features, the role relations, the equivalence relations, the shared ontologies.

We also expect that the database will be extended to many more languages. An important future task is here to maintain the framework so that the standardization effect will continue. Furthermore, it will be necessary to complete the specification to all parts-of-speech and to multi-word expressions. A lot of work is needed to further adapt the Inter-Lingual-Index to a truly universal and minimal index of meaning. Such a standardized list of meaning will be useful to compare and apply word-sense-disambiguation cross-linguistically and to standardize multilingual lexicons or translation systems.

The EuroWordNet exploitation consists of two components:

1.   distribution agreements with ELDA/ELRA
2.   the Global WordNet Association

All the builders of the wordnets have signed a bilateral distribution agreement with ELDA/ELRA. Each builder is thus responsible for the distribution of his/her own wordnet. They are also responsible for the clearing of claims and copy-rights of the data. The agreements are non-exclusive so that the builders/owners can also distribute their local wordnet independently. The agreements are the same for all the wordnets, where the fees depend on the number of synsets that have been licensed. The data are

publicly available through 4 types of licenses. The data have been delivered in the form of a general CD with all the public data and documentation and, for each language, a specific CD with documentation on the specific wordnet.

In addition to the direct exploitation of the EuroWordNet data, we planned to set up a Global Wordnet Association (GWA). We feel that the standardized framework of EuroWordNet should be continued after the project and extended to other (non-European) languages. This will stimulate further standardization of the wordnets and sharing of resources and technology across all associated languages.

The goal of GWA is to establish a network of excellence for maintaining, standardizing and interlinking wordnets for all languages in the world, likewise preparing the ground for the development of a world wide multilingual database with wordnets.

Appendices

Different ways of accessing the data structures in the EuroWordNet database

The EuroWordNet database has a special interface to match sets of synsets across wordnets. This can be down in several general ways:

1. multiple windows that expand separate wordnets and show the equivalence relations (see Figure 4)
2. looking up inter-lingual-index items (Explore ILI-records) which will give the associated synsets in each language (see Figure 5)
3. looking up Top-Concepts, which will give associated ILI-records (mostly Base Concepts) and the synsets in each language that are associated with these (see Figure 6)
4. looking up Domains, which will give associated ILI-records (mostly more specific concepts) and the synsets in each language that are associated with these (see Figure 7)
5. projecting a set of synsets in one language to a target language, via a selected set of equivalence relations (see Figure 8).
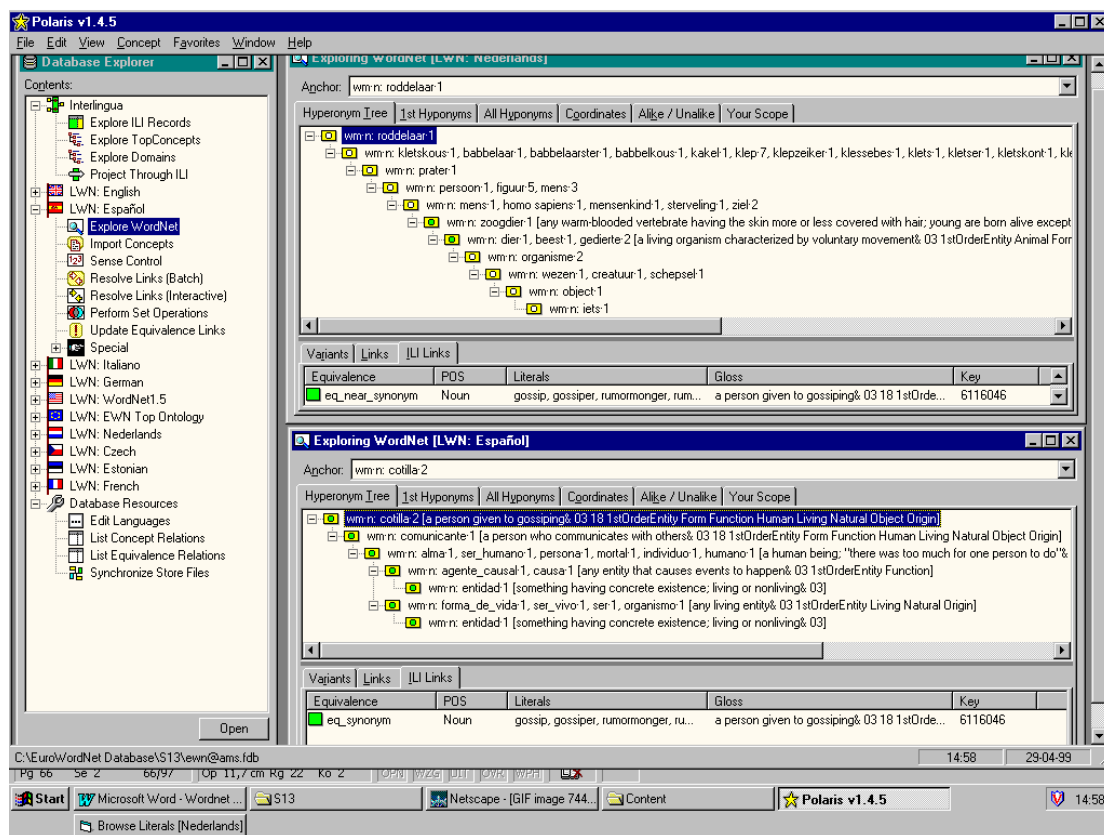

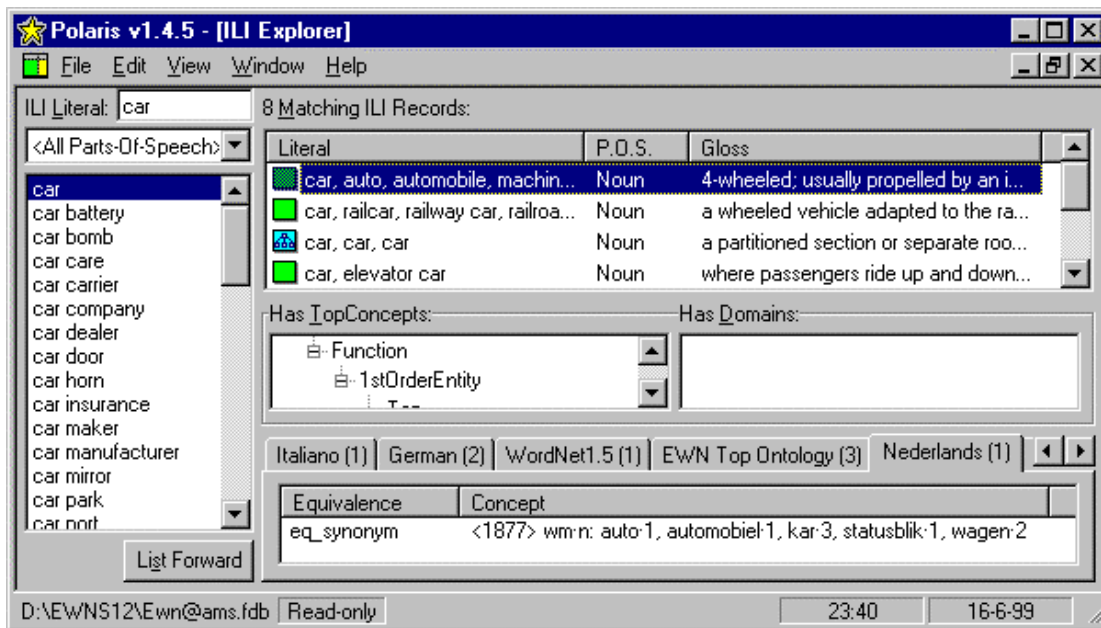*Figure 4: Accessing separate wordnets and their equivalence links*

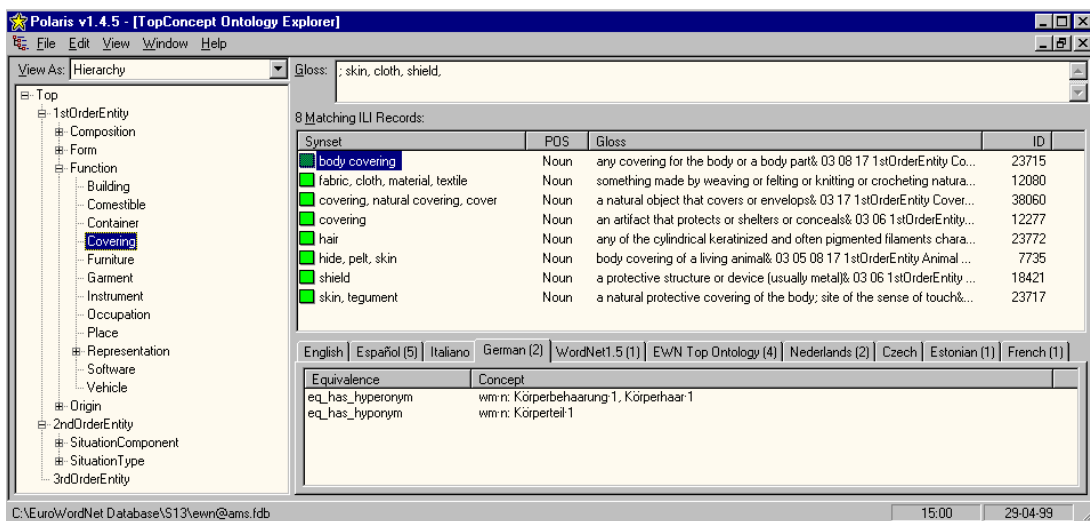*Figure 5: Accessing different wordnets via the Inter-Lingual-Index*

Figure 6: Accessing different wordnets via the Top-Ontology
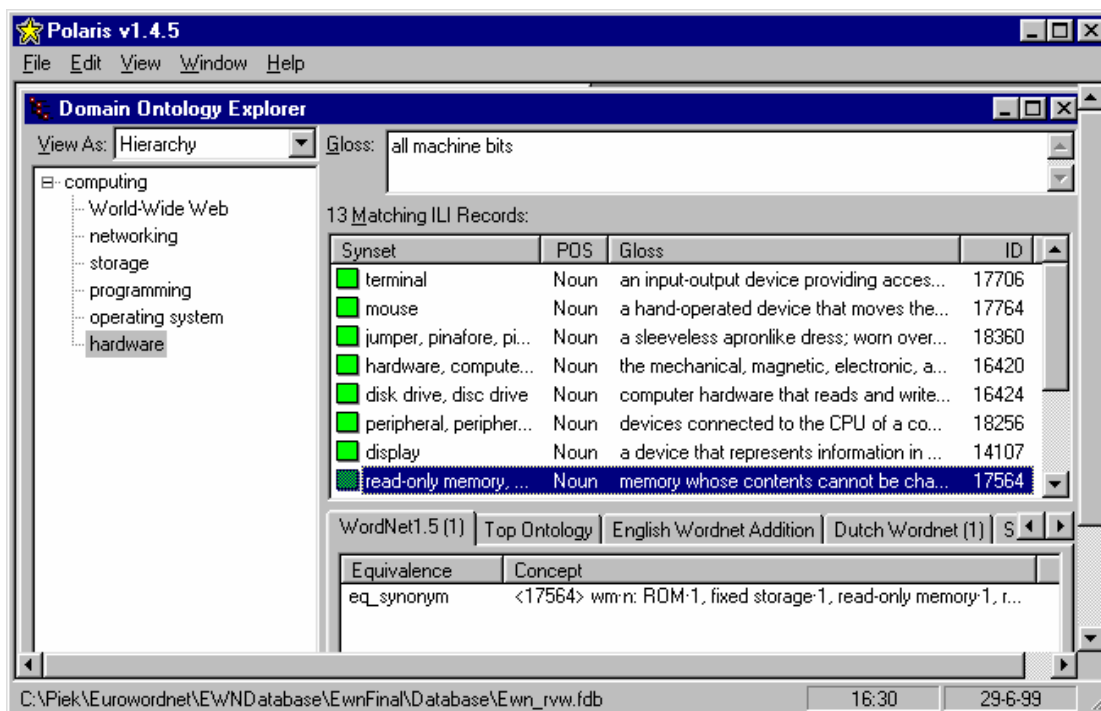
Figure 7: Accessing different wordnets via the Domain hierarchy
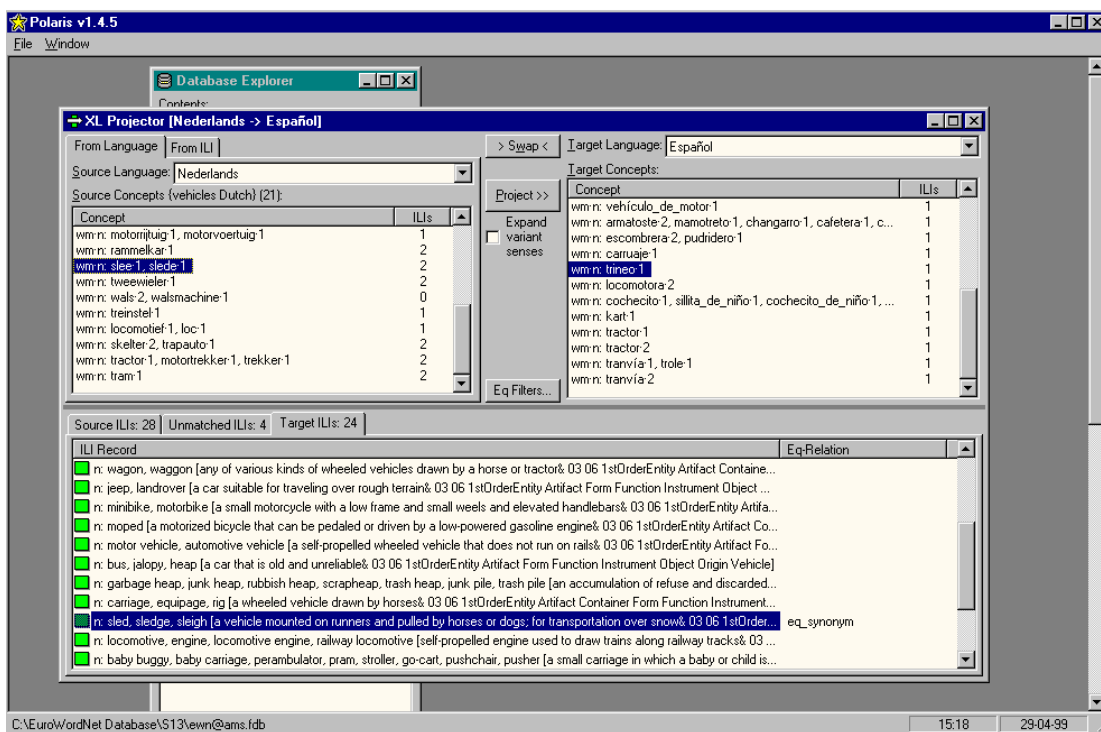
*Figure 8: Projecting Dutch "*vehicles*" (1 level) to the Spanish wordnet*

In the case of a projection, which is shown in Figure 8, a selection of synsets in a particular language (as shown in the left upper window for Dutch vehicles) is loaded and the desired types of equivalence mapping are selected. When a target language is chosen, the ILI-records that match the equivalence types are taken to generate the synsets in the target language also linked to them. The resulting set of target synsets is given in the right upper window, as is shown here for Spanish. The lower window gives, with different

TABs, the ILI-records that are linked in the source selection, the ILI-records that could not be matched and the records that are shared by the source and target.

The cross links can also be activated by double-clicking the synsets or the ILI-records. For example, double-clicking a ILI-record that is given as an equivalent for a synset in the language-specific explorer, will activate the ILI-explorer and from there it is possible to select a synset in another language.

List of public deliverables and reports

Deliverable Identifier	Title
D000	EuroWordNet project synopsis
D001	Bloksma, L., P. Díez-Orzas, P. Vossen 1996 User requirements and functional specification of the EuroWordNet project, EuroWordNet (LE-4003) deliverable D001, University of Amsterdam.
D005	Climent, S., Rodriguez, H., Gonzalo, J. (eds.) 1996 Definition of the links and subsets for nouns
D006	Alonge, A. (ed.) 1996 Definition of the links and subsets for verbs
D007	Diez-Orzas, P., P. Forest, M. Louw 1996 High-level Architecture of the EuroWordNet Databas
D008D012	Cuypers, Ilse, and Geert Adriaens 1997 EuroWordNet Viewer. EuroWordNet (LE-4003) Deliverable D008D012, University of Amsterdam.
D009	Vossen, P. 1997 Annual Report 1997. EuroWordNet (LE-4003) Deliverable D009, University of Amsterdam.
D010D011	Vossen, P. (ed.) 1997 Encoding the Semantic Relations for basic Nouns and Verbs. EuroWordNet (LE-4003) Deliverable D010D011. University of Amsterdam.
D013	Cuypers, I., A. Sánchez Valderrábanos, L. Schippers, G. Adriaens, M. Louw, P. Forest. 1997. Test specifications for EuroWordNet: internal data quality and application in multilingual information retrieval. EuroWordNet (LE-4003). Deliverable D013, University of Amsterdam.
D014D015	Vossen, P., L. Bloksma, S. Climent, M. Antonia Marti, G. Oreggioni, G. Escudero, G. Rigau, H. Rodriguez, A. Roventini, F. Bertagna, A. Alonge, C. Peters, W. Peters. 1998 The Restructured Core wordnets in EuroWordNet: Subset1. EuroWordNet (LE-4003) Deliverable D014D015, University of Amsterdam.
D017D034D036	Vossen, P., L. Bloksma, H. Rodriguez, S. Climent, N. Calzolari, A. Roventini, F. Bertagna, A. Alonge, W. Peters. 1997 The EuroWordNet Base Concepts and Top Ontology. EuroWordNet (LE 4003) Deliverable D017, D034, D036. University of Amsterdam
D021D025	Vossen, P., L. Bloksma, P. Boersma, F. Verdejo, J. Gonzalo, H. Rodriquez, G. Rigau, N. Calzolari, C. Peters, E. Picchi, S. Montemagni, W. Peters. 1998 EuroWordNet Tools and Resources Report. EuroWordNet (LE-4003) Deliverable D021D025, University of Amsterdam.
D023D024	M. Louw.1998. Polaris User's Guide.The EuroWordNet Database Editor. EuroWordNet (LE-4003), Deliverable D023D024, *Lernout & Hauspie - Antwerp, Belgium*
D026	Vossen, P. 1998 The EuroWordNet Annual Report 1998. EuroWordNet (LE-4003) Deliverable D026, University of Amsterdam.
D027D028	Vossen, P., L. Bloksma,S. Climent, M. A. Marti, M. Taule, J. Gonzalo, I. Chugur, M. F. Verdejo, G. Escudero, G. Rigau, H. Rodriguez, A. Alonge, F. Bertagna, R.

	Marinelli, A. Roventini, L. Tarasi. 1998. EuroWordNet Subset2 for Dutch, Spanish and Italian, EuroWordNet (LE-4003) Deliverable D027D028, University of Amsterdam.
D029D030	Vossen, P., S. Climent, M. A. Marti, M. Taule, J. Gonzalo, I. Chugur, M. F. Verdejo, G. Escudero, G. Rigau, H. Rodriguez, A. Alonge, F. Bertagna, R. Marinelli, A. Roventini, L. Tarasi. 1998. Comparison of the Final Wordnets Dutch, Spanish and Italian, EuroWordNet (LE-4003) Deliverable D029D030, University of Amsterdam.
D032D033	Vossen, P. (ed.) The Final wordnets for Dutch, Spanish, Italian and the English Addition, EuroWordNet (LE-4003) Deliverable D032D033, University of Amsterdam.
D032D033/2D014 Part A1	Vossen, P (ed.) EuroWordNet General Document. EuroWordNet (LE2-4003, LE4-8328), Part A, Final Document
D032D033/2D014 Part B1	Vossen, P (ed.) WordNet1.5 in EuroWordNet format
D032D033 Part B2	Vossen, P., L. Bloksma, P. Boersma The Dutch Wordnet, University of Amsterdam
D032D033 Part B3	Verdejo, M.Felisa The Spanish Wordnet, UNED, Madrid
D032D033 Part B4	Peters, W. The English Wordnet, University of Sheffield
D032D033 Part B5	Alonge, A., F. Bertagna, N. Calzolari, A. Roventini The Italian Wordnet
2D001	Vossen, P., C. Kunze, A. Wagner, D. Dutoit, K. Pala, P. Sevecek, K. Vider, L. Paldre, H. Orav, H. Õim. 1998 Revised Set of Common Base Concepts, EuroWordNet-2 (LE-8328), Deliverable 2D001, University of Amsterdam.
2D002	Dutoit, D., L. Catherin, C. Kunze, A. Wagner. 1998. Specification of German & French WNs. EuroWordNet (LE-8328) Deliverable: 2D002
2D003	Õim, H., K. Vider, L. Paldre, H. Orav, K. Pala, 1998. Specification of Czech and Estonian WNs. EuroWordNet (LE-8328) Deliverable: 2D003
2D004	Peters, W. and I. Peters. 1998. The Restructuring of the ILI. EuroWordNet (LE-8328) Deliverable 2D004
2D005	Wagner, A., D. Dutoit, L. Catherin. 1998. Tools & resources German & French WNs. EuroWordNet (LE-8328) Deliverable 2D005
2D006	Pala, Karel, Pavel Ševeček, Haldur Õim, Kadri Vider, Leho Paldre, Heili Orav, 1998. Tools & resources Estonian & Czech WNs . EuroWordNet (LE-8328) Deliverable 2D006
2D007	Kunze, C., A. Wagner, D. Dutoit, L. Catherin, K. Pala, P. Sevecek, K. Vider, L. Paldre, H. Orav, H. Oim. 1998. First WNs for BCs in French, German, Czech and Estonian. EuroWordNet (LE-8328) Deliverable 2D007
2D008	Laurent Catherin , Piek Vossen, Claudia Kunze, Andrea Wagner, Karel Pala, Kadri Vider, 1999 Compared and restructured wordnets for BCs in French, German, Czeck & Estonian, EuroWordNet (LE-8328) Deliverable 2D008

2D010	Piek Vossen, Laura Bloksma, Wim Peters, Claudia Kunze, Andreas Wagner, Karel Pala, Kadri Vider, Francesca Bertagna, 1999 Extending the Inter-Lingual-Index with new concepts, EuroWordNet (LE-8328) Deliverable 2D010
2D011D012	P. Vossen (ed.) 1999 Comparison of the German, French, Estonian and Czech wordnets
2D014 Part A	Claudia Kunze (eds) 1999 Final wordnets for German, French, Estonian and Czech, EuroWordNet (LE-8328) Deliverable 2D014
2D014 Part B1	Claudia Kunze and Andreas Wagner, 1999 The German Wordnet, EuroWordNet (LE-8328) Deliverable 2D014
2D014 Part B2	Laurent Catherin, 1999 The French Wordnet, EuroWordNet (LE-8328) Deliverable 2D014
2D014 Part B3	K. Vider, L. Paldre, H. Orav, H. Oim, 1999 The Estonian Wordnet, EuroWordNet (LE-8328) Deliverable 2D014
2D014 Part B4	K. Pala, P. Sevecek, 1999 The Czech Wordnet, EuroWordNet (LE-8328) Deliverable 2D014

Papers related to EuroWordNet

Code	Year	Reference
P001	1996	Castellón, I.,A. Martí, R. Morante, G. Vázquez 'Definición y Formalización de papeles temáticos en el marco del proyecto Pirápides' In: V. Simposio de Communicación Social. Centro de Lingüística Aplicada. Santiago de Cuba.
P002	1996	Agirre E. and G. Rigau 'Word Sense Disambiguation using conceptual density', Coling 1996.
P003	1996	Rigau G. and E. Agirre 'Linking Bilingual Dictionaries to WordNet'. Euralex 1996
P004	1996	Vossen, P. Right or wrong: combining lexical resources in the EuroWordNet project. In: M. Gellerstam, J. Jarborg, S. Malmgren, K. Noren, L. Rogstrom, C.R. Papmehl, Proceedings of Euralex-96, Goetheborg, 1996, 715-728.
P005	1996	Díez-Orzas, P. WordNet1.5 Contents and Statistics, Novell LTD, Internal Report, Antwerp.
P006	1996	Gilarranz, J., J. Gonzalo, and F. Verdejo. An Approach to Conceptual Text Retrieval Using the EuroWordNet Multilingual Semantic Database. Proceedings of AAAI-96 Spring Symposium Cross-Language Text and Speech retrieval.
P007	1996	Cuypers I., P. Díez-Orzas, P. Forest, M. Louw,L. Schipper, P. Vossen The EuroWordNet Viewer, Novell LTD, Internal Paper, Antwerp
P008	1996	Cuypers and Díez-Orzas The Manual of the Novell ConceptNet Management System & Tookit 1.0
P009	1997	Hamp, B. & H. Feldweg 1997. GermaNet - a lexical-semantic net for German. In: P. Vossen, N. Calzolari, G. Adriaens, A. Sanfilippo & Y. Wilcks (Eds.). Proceedings of the ACL/EACL-97 workshop on Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications. Madrid, 7--12 July, 1997, pp. 9-15.
P010	1997	Feldweg, H. 1997. GermaNet - ein lexikalisch-semantisches Netz fuer das Deutsche. Vortrag auf dem Workshop: Lexikalische Semantik aus kognitiver Sicht. Perspektiven im Spannungsfeld linguistischer und psychologischer Modellierungen. Osnabrueck, 02--04 October, 1997. to appear: Gunter Narr Verlag: Tuebingen.
P011	1997	Louw, M. EuroWordNet import specifications, Novell LTD, Internal Paper, Antwerp
P012	1997	Vossen, P. EuroWordNet: a multilingual database for information retrieval, In: Proceedings of the DELOS workshop on Cross-language Information Retrieval, March 5-7, 1997, Zurich.
P013	1997	Vossen, P., P. Diez-Orzas, W. Peters The Multilingual Design of EuroWordNet. Published in: P. Vossen, N. Calzolari, G. Adriaens, A. Sanfilippo, Y. Wilks (eds.) Proceedings of the ACL/EACL-97 workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications, Madrid, July 12th, 1997.
P014	1997	Vossen, P., W. Peters, P. Díez-Orzas, The Multilingual design of the EuroWordNet Database, In: Kavi Mahesh (ed.) Ontologies and multilingual NLP, Proceedings of workshop at IJCAI-97, Nagoya, Japan, August 23-29.
P015	1998	J. Atserias, I. Castellón, M. Civit. 1998. "Syntactic Parsing of the Unrestricted Spanish Text". Proceedings of The 1st International Conference on Language Resources Association, LREC'98 , Granada, 1998
P016	1998	J. Atserias, J. Carmona, I. Castellón, S. Cervell, M. Civit, L.Márquez, M.A. Martí, L. Padró, R. Placer & M. Taulé.

		1998. "Morphosyntactic Analysis and Parsing of Unrestricted Spanish Text" Proceedings of 1st International Conference on Language Resources and Evaluation, LREC'98. Granada 1998, Spain.
P017	1998	L. Benítez, S. Cervell, G. Escudero, M. López, G. Rigau & M. Taulé. 1998. "Methods and Tools for Building the Catalan WordNet", Proceedings of the ELRA Workshop on Language Resources for European Minority Languages, Granada,Spain.1998.
P018	1998	J. Carmona, S. Cervell, J. Atserias, J. Carmona, I. Castellón, S. Cervell, M. Civit, L. Márquez, M.A. Martí, L. Padró, R. Placer, H. Rodríguez, M. Taulé & J. Turmo. 1998, An Environment for Morphosyntactic Processing of Unrestricted Spanish Text". Proceedings of 1st International Conference on Language Resources and Evaluation, LREC'98. Granada, Spain 1998.
P019	1998	B. Dorr, M.A.Martí & I. Castellón. 1998. Spanish EuroWordNet and LCS-Based Interlingual MT". Proceedings of 1st International Conference. European Resources and Evaluation, LREC'98, Granada, 1998
P020	1998	X. Farreres, G. Rigau & H. Rodríguez. 1998 "Using WordNet for Building WordNets". In Proceedings of COLING-ACL Workshop "Usage of in Natural Language Processing Systems". Montreal, Canada. 1998.
P021	1998	A. Fernández Montraveta, M. A. Martí, G. Vázquez, I. Castellón y R. Morante 1998. "Criterios de clasificación de las estructuras eventivas: el cambio y la transferencia", RESLA (in press).1998.
P022	1998	A. Fernández Montraveta y M. A. Martí Antonín 1998. "Hacia una propuesta de desajustes verbales", RESLA (in press). 1998.
P023	1998	J. Gonzalo, F. Verdejo, I. Chugur and J. Cigarrán 1998 "Indexing with WordNet synsets can improve text retrieval", Proceedings of the ACL/COLING Workshop on Usage of WordNet for Natural Language Processing. Sanda Harabagiu (Ed). 1998.
P024	1998	J. Gonzalo, F. Verdejo, I. Chugur, Fernando López, Anselmo Peñas. 1998. "Extracción de relaciones semánticas entre nombres y verbos en EuroWordNet". Revista de la SEPLN, nº23,1998.
P025	1998	L. Márquez, L. Padró & H. Rodríguez. 1998. "Etiquetado Morfosintáctico de Corpus Textuales". Congreso Anual de la Asociacion Española de Linguística Aplicada, AESLA'98. Logroño, 1998.
P026	1998	L. Márquez & H. Rodríguez. 1998. "Part-of-Speech-Tagging Using Decision Trees" In "Proceedings of the 10th European Conference on Machine Learning". (ECML'98). Chemnitz, Germany. April 1998.
P027	1998	R. Morante, I. Castellón & G. Vázquez. 1998. "Los verbos de trayectoria", Revista de la SEPLN,nº23. 1998.
P028	1998	L. Padró & L. Márquez. 1998. "On the Evaluation and Comparison of Taggers: The Effect of Noise in Training Corpora". Proceedings of Joint COLING/ACL, 1998. Montreal, Canada.
P029	1998	L. Padró, L. Márquez & H. Rodríguez. 1998. "Improving Tagging Accuracy by Using Voting Taggers". In Proceedings of International Conference on Natural Language Processing and Industrial Applications, NLP+IA/TAL+AI '98. Moncton, New Brunswick, Canada, 1998.
P030	1998	G. Rigau, H. Rodríguez & E. Agirre. 1998. "Building Accurate Semantic Taxonomies from Monolingual MRDs", in proceedings of the 17th International Conference on Computational Linguistics and 36th Annual Meeting of the Association for Linguistics (COLING-ACL'98). Montreal, Canada. 1998.
P031	1998	Roventini A., Bertagna F., Peters C., Calzolari N., 1998. Building the Italian Component of EuroWordNet: a Langage-specific Perspective, in:

		Thierry Fontenelle et al. (eds), Euralex'98 Proceedings, Vol. I, pp. 249-258.
P032	1998	Alonge A., 1998. Encoding data on lexicalisation of semantic components: the EuroWordNet relational model, in: Thierry Fontenelle et al. (eds), Euralex'98 Proceedings, Vol. I, pp.45-56.
P033	1998	Peters C., Roventini A., Marinai E., Calzolari N., 1998. Making the Right Connections: Mapping between Italian and English Lexical Data in EuroWordNet, in: ALLC/ACH '98 Conference Abstracts, Lajos Kossuth University, July, 5-10 1998, pp. 130 - 133.
P034	1998	Roventini A.,Calzolari N., Peters C., Bertagna F., 1998. Building a Semantic Network for Italian using Existing Lexical Resources, in Antonio Rubio et al. Editors: First International Conference on Language Resources & Evaluation Proceedings, Granada 28-30 May 1998, pp. 377- 384.
P035	1998	Alonge A. 1998. Data on Verb Semantics in the EuroWordNet Database, in Antonio Rubio et al. Editors: First International Conference on Language Resources & Evaluation Proceedings, Granada 28-30 May 1998, pp.385-392.
P036	1998	E. Hinrichs, C. Kunze & A. Wagner 1998. Integrating GermaNet into EuroWordNet, a multilingual lexical-semantic database. German-Japanese Workshop on "Lexical Resources for Information Retrieval". Stuttgart, 15--16 October, 1998. In: B. Schroeder (ed): Sprache und Datenverarbeitung. Vol. 2/1999.
P037	1998	Gonzalo, J., F. Verdejo, C. Peters, N. Calzolari., Applying EuroWordNet to Multilingual Text Retrieval, in: Computer and the Humanities, Special Issue on EuroWordNet.
P038	1998	Louw, M. The ILI import specifications, Novell LTD, Internal Paper, Antwerp
P039	1998	Vossen, P. 1998. EuroWordNet: Building a Multilingual Database with Wordnets for European Languages. In: K. Choukri, D. Fry, M. Nilsson (eds), The ELRA Newsletter, Vol3, n1, 19998. ISSN: 1026-8200.
P040	1998	Vossen P. and L. Bloksma, 1998. Categories and classifications in EuroWordNet. In: A. Rubio, N. Gallardo, R. Catro and A. Tejada (ed) Proceedings of First International Conference on Language Resources and Evaluation, Granada, 28-30 May 1998. 399-408
P041	1998	Peters, W., I. Peters, and P. Vossen, 1998. The Reduction of Semantic Ambiguity in Linguistic Resources. In: A. Rubio, N. Gallardo, R. Catro and A. Tejada (ed) Proceedings of First International Conference on Language Resources and Evaluation, Granada, 28-30 May 1998. 409-416
P042	1998	Vossen, P., 1998. EuroWordNet: Linguistic Ontologies in a Multilingual Database. In: D. Vervenne, F. Vandamme (eds), "Towards advanced document management: the creation and maintenance of electronic thesauri", CC AI, Communication and Cognition for Artificial Intelligence, Special Issue, Volume 15, Numbers 1-2, 1998. 37-80.
P043	1998	Vossen, P. 1998.Vossen, P. Introduction to EuroWordNet. In: Nancy Ide, Daniel Greenstein, Piek Vossen (eds), Special Issue on EuroWordNet. Computers and the Humanities, Volume 32, Nos. 2-3 1998. 73-89.
P044	1998	Alonge, A., N. Calzolari, P. Vossen, L. Bloksma, I. Castellon, T. Marti, W. Peters, 1998 The Linguistic Design of the EuroWordNet Database. In: Nancy Ide, Daniel Greenstein, Piek Vossen (eds), Special Issue on EuroWordNet. Computers and the Humanities, Volume 32, Nos. 2-3 1998. 91-115.
P045	1998	Rodriquez, H., S. Climent, P. Vossen, L. Bloksma; A. Roventini, F. Bertagna, A. Alonge, W. Peters,

		1998. The Top-Down Strategy for Building EuroWordNet: Vocabulary Coverage, Base Concepts and Top Ontology. In: Nancy Ide, Daniel Greenstein, Piek Vossen (eds), Special Issue on EuroWordNet. Computers and the Humanities, Volume 32, Nos. 2-3 1998. 117-152.
P046	1998	Vossen, P., L. Bloksma, C. Peters, A. Alonge, A. Roventini, E. Marinai, I. Castellon, T. Marti, G. Rigau, 1998. Compatibility in Interpretation of Relations in EuroWordNet. In: Nancy Ide, Daniel Greenstein, Piek Vossen (eds), Special Issue on EuroWordNet. Computers and the Humanities, Volume 32, Nos. 2-3 1998. 153-184.
P047	1998	Peters, W., P. Vossen, P. Diez-Orzas, G. Adriaens, 1998. Cross-linguistic Alignment of Wordnets with an Inter-Lingual-Index. In: Nancy Ide, Daniel Greenstein, Piek Vossen (eds), Special Issue on EuroWordNet. Computers and the Humanities, Volume 32, Nos. 2-3 1998. 221-251.
P048	1998	Sanfilippo A., N. Calzolari, S. Ananiadou, R. Gaizauskas, P. Saint-Dizier, P. Vossen (eds). 1998. EAGLES, Prelimenary Recommendations on Semantic Encoding. Interim Report. February, 1988.
P049	1998	Diez-Orzas, P. L. and A. Alonge, 1998 'Exploiting data from the EuroWordNet database for Industrial Applications', in Antonio Rubio et al. (eds.) First International Conference on Language Resources & Evaluation Proceedings, Granada 28-30 May 1998, pp. 857-64.
P050	1998	Climent S. 1998 Individuación e Información Parte-Todo. Representación para el procesamiento computacional del lenguaje. PhD thesis. Facultad de Filosofía, Universidad de Barcelona
P051	1999	Gonzalo, J., F. Verdejo and I. Chugur (1999). Using EuroWordNet in a Concept-Based. Approach to Cross-Language Text Retrieval, in Applied Artificial Intelligence, Special on Multilinguality in the Software Industry: the AI contribution.
P052	1999	Piek Vossen, Wim Peters and Julio Gonzalo (1999). Towards a Universal Index of Meaning. In proceedings of the ACL-99 Siglex workshop, University of Maryland.
P053	2000	Kunze, C. & A. Wagner: Anwendungsperspektiven des GermsNet, eines lexikalisch-semantischen Netzes für das Deutsche. In: Lemberg, I. & A. Storrer & B. Schroeder (eds.): Probleme und Perspektiven computergestützter Lexikographie. Tübingen:Niemeyer.
P054	2000	Kunze, C. (1999): Semantics of Verbs within GermaNet and EuroWordNet. In: Kordoni, E.(ed.): Lexical Semantics and Linking in Constrained-Based Theories. Workshop Proceedings of 11th ESSLLI. Utrecht, 16th --20th August, 1999.

Books related to EuroWordNet

1997	Vossen, P., N. Calzolari, G. Adriaens, A. Sanfilippo, Y. Wilks (eds) 1997 Proceedings of the ACL/EACL-97 workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications, Madrid, July 12th, 1997.
1998	Vossen, P. (eds) 1998 EuroWordNet: A Multilingual Database with Lexical Semantic Networks, Kluwer Academic Publishers, Dordrecht

EuroWordNet User-Group

Organisation/ Department	First Name	Last Name	City	Country	Application area
Bibliograf	Juan P.	Acordagoic oechea	Barcelona	Spain	(electronic) dictionaries
Cambridge Language Services Ltd.	Susan	Allen-Mills	Cambridge	Great Britain	electronic dictionaries, languae learning tools
Cap Volmac	Pim	van der Eijk	Utrecht	The Netherlands	authoring tools, grammar checkers
Carnegie Mellon Univ Dept of CS	Doug	Beeferman	Pittsburgh	USA	
Centre of Computational Linguistics	Ruta	Marcinkevi ciene	Kaunas	Lithuania	lithuanian wordnet
Centro de Investigacions Linguisticas e Literarias	Fernando	Magan	SANTIAG O DE COMPOST ELA	Spain	
Centro de Linguistica da Universidade de Lisboa	Amalia	Mendes	Lisboa	Portugal	
Copenhagen Buzines School dept of computational linguistics	Sabine	Kirchmeier Anderson	Frederiksbe rg	Denmark	teaching and research
Datamat	Gian Franco	Abbrescia	Rome	Italy	information retrieval, document processing
Debili (CNRS)		Debili	Fontenay aux Roses	France	traitement du francais et de l'arabe- indexation - Cross language
Dept. Informatica Universidad de Jaen	Laurena	Alfonso Urena	Jaen	Spain	
Dept. of Computerscience, University of science 7	Wu	Dekai	Hong Kong	Hong Kong	
Dept. of Swedisch University of Goteborg	Maria	Toporowsk a Gronostaj	Goteborg	Sweden	
DFKI	Paul	Buitelaar	Saarbrueck en	Deutschland	resources in various applications, information extraction and concept-based engines; focus ons EN, DU, Ger, It
EBSCO Subscription Service	Cristina	de la Pena	Madrid	Spain	products for automated library systems
ERLI	Antoine	Ogonowski	Charenton cedex	France	software producer
Facultad de Traduccion e interpretacion	Maite	Saveedra Valle	Valdepenas	Spain	

Organisation/ Department	First Name	Last Name	City	Country	Application area
Francetelecom - cnet	Isabelle	Metayer	Lamion	France	operateur de telecommunication
I.A.I	Paul	Schmidt	Saarbrucken	Deutschland	
IBM	Louis	Sopena	Madrid	Spain	
IMC Bureau	Frits	van Latum	Rotterdam	The Netherlands	
Incyta s.l.	Juan A.	Alonso	Cornella	Spain	machine translation
Infolab, Tilburg University	Dr. Hans	Weigand	Tilburg	The Netherlands	The task is to provide a multilingual Lexicon Management System to support the automatic parsing and subject identification.
Institute for Language and Speech Processing	Maria	Gavrilidou	Athene	Greece	
Instytut Informatyki Uniwersytetu Warszawskiego	Janusz	Bien	Warszawa	Poland	
IRIT-CNRS	Patrick	Saint-Dizier	Toulouse cedex	France	
John's Hopkins University dpt of Computer Science	David	Yarowsky	Baltimore	USA	NLP research group
Kenniscentrum Cibit	Cor	Baars	Utrecht	The Netherlands	knowledge engineering
Komet Multilingual Text Generation	Dr. John A.	Bateman	Darmstadt	Germany	cross linguistic ontologies
Lexicon Research Group, dpt. Ling.Sci University o	Elizabeth	Hallum	Reading	UK	emperical research ons adult language acquisition: need resources to predict cross-linguistic differences/problems
Lingsoft Inc	Markku	Norberg	Helsinki	Finland	sofware company
Localisation Resouces Centre, UCD	Reinhard	Schaeler	Dublin-4	Ireland	Focus point and research support centre for the localisation industry
LOGOS-GROUP	Cinzia	Bazzani	Modena	Italy	technical translations/ desktop publishing
Lund University Department of Linguistics & Phonetics	Åke	Viberg	Lund	Sweden	
Medialab b.v.	M.M.	Chanowski	Schellinkhout	Nederland	
National Security Agency, RSL	Charles	Wayne	Maryland	USA	research

Organisation/ Department	First Name	Last Name	City	Country	Application area
NCSR - Demokritos - Inst. of Informatics 2 Telecom	Vangelis	Karkaletsis	Athene	Greece	research organisation
Office Line Engineering NV	Werner	Ceusters	Zonnegem	Belgium	
Oxford Logic, Inc.	Frank	White	Edinburgh	Scotland	
Princeton University/ Cognitive Science	Christiane	Fellbaum	Princeton	USA	American wordnet
RKD	J.H.E.	van der Starre	Den Haag	The Netherlands	information retrieval, electronic libraries
SENA	Perikles	Tsahageas	Filothei	Greece	information retrieval
Sun Microsystems Laboratories	Philip	Resnik	Chelmsford	USA	knowledge engineering, Information Retrieval
Talana University Paris	A.	Abeille	Paris	France	research in French NLP
TNT Express Worldwide			Amsterdam	The Netherlands	courier
Univ. do Minho, Dept. de Informatica	Jose Joao	Diaz de Almeida	Braga	Portugal	
Universidad Alfonso X El Sabio	Pedro Luis	Diez-Orzas	Madrid	Spain	
Universidad Alfonso X El Sabio	Aurora	Martmn de Santa Olalla	Madrid	Spain	machine translation, corpus linguistics
Universidad Autonoma de Barcelona Laboratorio de Lingüística Informática	Carlos	Subirats-Rüggeberg	Bellaterra	Spain	
Universidad Autonoma de Madrid	Dr. Antonio	Moreno Sandoval	Madrid	Spain	information extraction, machine translation
Universidad Computense Madrid	Isabel	Santos	Madrid	Spain	
UNIVERSIDAD EUROPEA DE MADRID	Manual	Buenaga	Madrid	Spain	
Universidade de Lisboa Fac de Letras-dpt de Lingüística	Palmira	Marrafa	Lisboa Codex	Portugal	
Universidade Nova de Lisboa	Gael	Hany Dias	Lisboa	Portugal	Acquisition of lexical units
Universitat Leipzig Institut fur Informatik	Uwe	Quasthoff	Leipzig	Deutschland	
University of Athens	Stelios	Piperidis	Athens	Greece	Greek wordnet
University of Dundee, Microcentre	Stefan	Langer	Dundee	Great Britain	research on applications of WordNet for message retrieval

Organisation/ Department	First Name	Last Name	City	Country	Application area
University of Euskal Herriko/ Informatika	Arantza	Diaz de Llarraza	Donostia	Basque Spain	Basque wordnet
University of Goteborg	J.	Jarborg	Goteborg	Sweden	Swedish wordnet
University of Heidelberg, Institute for Computer Linguistics	Peter	Hellwig	Heidelberg	Germany	German wordnet
University of Maryland	Eric Lee	Peterson	Alexandria	USA	
University of Sunderland	Theoktisti	Zervaki	Sunderland	UK	
University of Sussex	John	Carrou	Brighton	UK	NLP academic basic and applied research
USC Information Sciences Institute	Eduard	Hovy	Admiralty Way	USA	
Van Dale Lexicografie B.V.	Margreet A.	Moerland	Utrecht	The Netherlands	electronic dictionaries, language generation
VPRO	Bruno	Felix	Hilversum	The Netherlands	information retrieval, Internet services
Vrije Universiteit	Hans	Burg	Amsterdam	The Netherlands	project Grammalizer