WordNet, EuroWordNet and Global WordNet.

Piek Vossen.

Piek.Vossen@irion.nl

1 WordNet

In 1978, George Miller started the development of a database with conceptual relations, as an implementation of a model of the mental lexicon. The database, called WordNet, was organized around the notion of a synset between which semantic relations are expressed. 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.

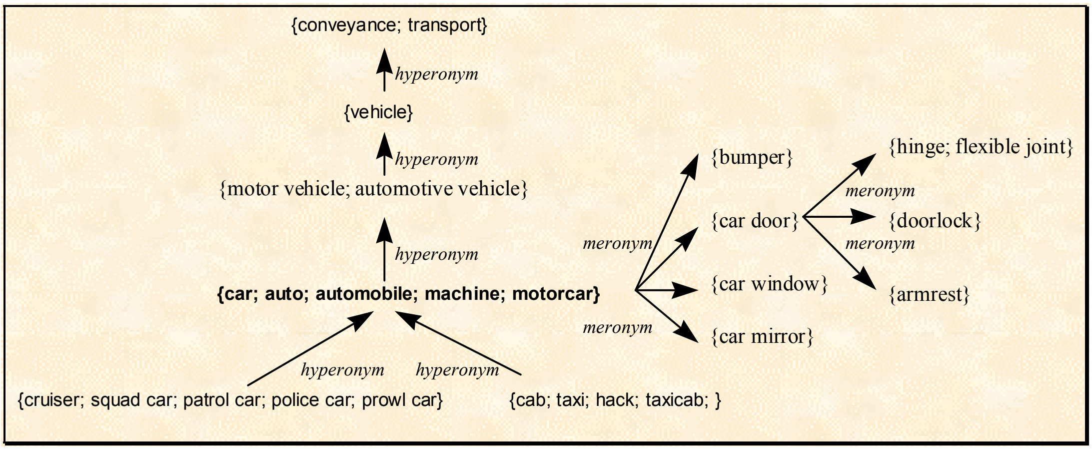

Figure 1: Synsets related to "car" in its first sense in WordNet1.5.

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

- A more general concept 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 word meanings in a language can be interconnected, constituting a huge network or wordnet. Such a wordnet can be used for making semantic inferences (what things can be used as *vehicles*), for finding alternative expressions or wordings (what words can refer to *vehicles*), or for simply expanding words to sets of semantically related or close words, in e.g. 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. In Fellbaum (1998), a detailed description is given of the history, background and characteristics of the Princeton WordNet.

WordNet now includes around 100,000 concepts and 150,000 word meanings. It has been used widely in many Natural Language Applications all over the world and has stimulated many new areas of research in the last 10 years.

2 EuroWordNet

WordNet only covered English and there was a growing need to develop similar resources for other languages. Specifically, a resource was needed that would relate and unite wordnets in different languages in a single multilingual lexical resource. In 1996, EuroWordNet began as an EU project, with the goal of developing wordnets for Dutch, Spanish and Italian, and to link these wordnets to the English wordnet in a multilingual database. In 1997, the project was extended to include German, French, Czech and Estonian. EuroWordNet (http://www.hum.uva.nl/~ewn) was completed at the end of 1999.

While the EuroWordNet database is in its core compatible with the English WordNet, there have been numerous additions and extensions to the relations and objects.

2.1 The overall design of the database

The design of 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 were 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;

The most important difference of EuroWordNet with respect to WordNet is its multilinguality, which also raises some fundamental questions with respect to the status of the monolingual information in the wordnets. In principle, multilinguality is achieved by adding an equivalence relation for each synset in a language to the closest synset in WordNet1.5. Synsets linked to the same WordNet1.5 synset are supposed to be equivalent or close in meaning and can then be compared. However, what should be done with differences across the wordnets? If 'equivalent' words are related in different ways in the different resources, we have to make a decision about the legitimacy of these differences. For example, in the Dutch wordnet we see that *hond* (dog) is both classified as *huisdier* (pet) and *zoogdier* (mammal). However, there is no equivalent for *pet* in Italian, and the Italian *cane*, which is linked to the same synset *dog*, is only classified as a *mammal* in the Italian wordnet.

EuroWordNet took the position that it must be possible to reflect such differences in lexical semantic relations. The wordnets are seen as linguistic ontologies rather than ontologies for making inferences only. In an inference-based ontology it may be the case that a particular level or structuring is required to achieve a better control or performance, or a more compact and coherent structure. For this purpose it may be necessary to introduce artificial levels for concepts, which are not lexicalized in a language (e.g. *natural object*, *external body parts*), or it may be necessary to neglect levels (e.g. *watchdog* or *silverware*) that are lexicalized but not relevant for the purpose of the ontology. A linguistic ontology, on the other hand, exactly reflects the lexicalization and the relations between the words in a language. It is a "wordnet" in the true sense of the word and therefore captures valuable information about conceptualizations that are lexicalized in a language: what is the available fund of words and expressions in a language. In addition to the theoretical motivation there is also a practical motivation for considering the wordnets as autonomous networks. To be more cost-effective, they have (as far as possible) been derived from existing resources, databases and tools. Each sites therefore had a different starting point for building their local wordnet, making it necessary to allow for a maximum of flexibility in producing the wordnets and structures.

To be able to maintain the language-specific structures and to allow for the separate development of independent resources, 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 are made explicit in the so-called Inter-Lingual-Index (ILI). Each synset in the monolingual wordnets has at least one equivalence relation with a record in this ILI, either directly or indirectly via other related synsets. 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*.

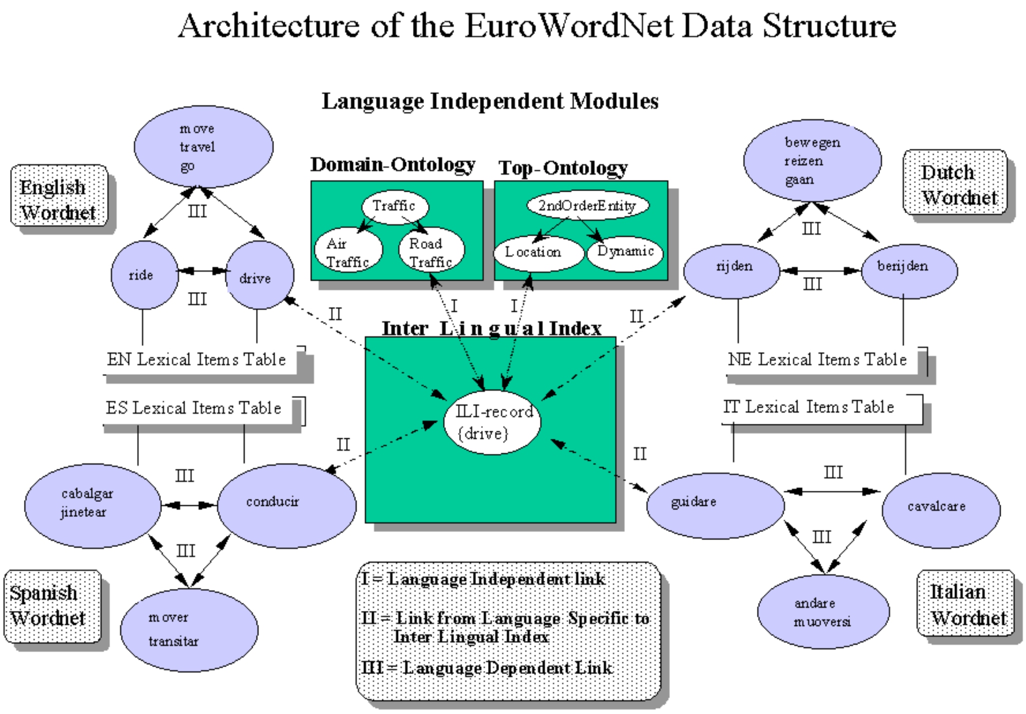

Figure 2. The global architecture of the EuroWordNet database.

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 ILI consists of a list of so-called ILI-records (ILIRs) which are related to word-meanings in the language-internal modules, (possibly) to one or more Top Concepts and (possibly) to domains. 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. Once the wordnets are properly linked to the ILI, 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.

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. The development of a complete language-neutral ontology is considered to be too complex and time-consuming given the limitations of the project. As an unstructured list, there is no need to discuss changes or updates to the index from a many-to-many perspective. Note that it will nevertheless be possible to indirectly see a structuring of a set of ILI-records by viewing the language-internal relations of the language-specific concepts that are related to the set of ILI-records. Since WordNet1.5 is linked to the index in the same way as any of the other wordnets, it is still possible to recover the original internal organization of the synsets in terms of the semantic relations in WordNet1.5.

The advantages of an interlingua such as the Inter-Lingual-Index are well-known in MT translation (Copeland et al. 1991, Nirenburg 1989):

1. It is not necessary to specify many-to-many equivalence relations between each language-pair and to have consensus across all the groups on the equivalence relations: each group only considers the equivalence relations to the Index.
2. New languages can be added without having to reconsider the equivalence relations for the other languages (e.g. BalkaNet, Stamou et al 2002a).
3. It is possible to adapt the Inter-Lingual-Index as a central resource to make the matching more efficient or precise. Updates can be made relatively easy because the ILI lacks any further structure.

Some language-independent structuring of the ILI is nevertheless provided by two separate ontologies, which may be 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. 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.

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. So far we have only included domain labels for computer terminology in EuroWordNet. However, users of the database can freely add domain labels to the ILI or adjust the top ontology without having to access or consider the language-internal relations of each wordnet (Magnini et al. 2001). In the same way, it is possible to extend the database with terminology or other ontologies provided that they are specified according to the EuroWordNet format and include a proper linking to the ILI, e.g. EuroTerm (Stamou et al. 2002b).

2.2 The language-internal relations

The most important relations of the Princeton wordnet have also been maintained in EuroWordNet. The relations have been extended in the following ways:

1. Relations can have features.
2. Existing relations have been broadened.
3. New relations have been added.

Relation features differentiate the implicational effect of relations: conjunction, disjunction, factivity, reversal, and negation of relations. For example, conjunction of relations would apply to all the parts (walls, windows, doors, roof) that together or conjunctively make up a whole (a house), whereas certain parts such as a door can be parts of different types of things (rooms, houses, vehicles) but only disjunctively. Similarly, factivity differentiates causal relations such as “to-kill-causes-to-die” from a non-factive relation such as “to-search-may-cause-to-find” in terms of necessity of the implication.

In WordNet, nouns, verbs and adjectives form separate sub-networks that are not interrelated. This strict separation between the parts of speech has been abandoned in EuroWordNet. For three major reasons:

- Languages can have lexicalizations of the same concept with different parts of speech, e.g the verb like in English matches with the adjective “aardig” in Dutch.
- There are many variants in languages of words that can refer to the same concepts but differ in part of speech, e.g. “to depart” and “departure”
- There are many salient and explicit semantic relations across part of speech of a thematic nature, e.g “teacher” and “to teach” or “scissor” and “to cut”.

This has led to the introduction of many new relations and broadening of some existing relations. For example, the causative relation was broadened to other parts of speech: “to redden” (verb) causes a state of “red” (adjective). Examples of new relations are: cross-part-of-speech synonymy between “depart” and “departure”, and role relations such as agent, instrument, patient and location.

The part-of-speech networks are thus highly connected in EuroWordNet. This gives many new possibilities to use the networks for word-sense-disambiguation, language generation and lexical expansion in information retrieval. It also gives a more precise framework to describe the exact relational and semantic structure of the vocabularies in certain areas of the vocabulary. Causal verbs are for example hardly distinguishable by just synonymy and hyponymy relations (e.g. “to crush”, “to smash”, “to smoothen”, “to spread” are all related to “change”) but still have a clearly distinguishable pattern of other relations; e.g. the resulting state of the process, the manner in which it is achieved, the instrument that is used, etc.

Table 1: Overview of language-internal relations

Relations	Examples
NEAR_SYNONYM	apparatus_N<=>machine_N<=>device_N
XPOS_NEAR_SYNONYM	move_V<=>movement_N
HAS_HYPERONYM/ HYPONYM	car_N<=>vehicle_N; walk_V<=>move_V
HAS_XPOS_HYPERONYM/ HYPONYM	hate_V<=>emotion_N
HAS_HOLONYM/ MERONYM (RELATION SUBTYPES: PART, MEMBER, PORTION, MADEOF, LOCATION)	head_N<=> nose_N ;player_N<=> team_N ; liquid_N<=> drop_N; stick_N<=> wood_N;centre_N<=> city_N
ANTONYM	do_V<=>undo_V
NEAR_ANTONYM	open_V<=>close_V
XPOS_NEAR_ANTONYM	sleep_V <=>sleep_A
CAUSES/ IS_CAUSED_BY	kill_N<=>die_N; wake up_N<=>awake_A; whiten_V<=>white_A; search_V<=>find_V; send_V<=>receive_V
HAS_SUBEVENT/ IS_SUBEVENT_OF	snore_V<=>sleep_V; buy_V<=>pay_V
ROLE/ INVOLVED (RELATION SUBTYPES: AGENT, PATIENT, INSTRUMENT, LOCATION, DIRECTION, SOURCE, TARGET, RESULT)	teacher_N<=>teach_V; patient_N<=>cure_V; crystalize_V<=>crystal_N; teach_V<=>school_N; land_V<=>ground_N; take off_V<=>sky_N
CO_ROLE (RELATION SUBTYPES: AGENT-PATIENT, AGENT-INTRUMENT, AGENT-RESULT, PATIENT-INSTRUMENT, PATIENT-RESULT, INSTRUMENT-RESULT)	teacher_N<=>student_N; docter_N<=>patient_N; hammer_N<=>nail_N; ice-saw_N<=>ice_N; guitar player_N<=>guitar_N; violist_N<=>violin_N; camera_N<=>picture_N;
IN_MANNER/ MANNER_OF	rush_V<=>quickly_A; slurp_V<=>noisely_A
BE_IN_STATE/ STATE_OF	poor_N<=>poor_A

Each relation in EuroWordNet is defined using a diagnostic frame (Cruse 1987). Different frames have been made for each language. A more detailed account of all the relations is given in Vossen 1999.

2.3 The multilingual relations

Each synset in a language can have one or more relation to the Inter-Lingual-Index (ILI). Different types of equivalence relations are allowed to handle various kinds of mismatches across languages. The most important relations are:

EQ_SYNONYM: there is a direct match between a synset and an ILI-record, e.g *diventare*$_V$ (Italian) and *become*$_V$ (English).

EQ_NEAR_SYNONYM: a synset matches multiple ILI-records simultaneously, e.g. *schoonmaken*$_V$ (Dutch) and various senses of *clean*$_V$ (English)

EQ_NEAR_SYNONYM: multiple near_synonym synsets match multiple near_synonym records in English, e.g. *apparaat, machine, toestel* (Dutch) and *apparatus, machine, tool* (English).

HAS_EQ_HYPERONYM: a synset is more specific than any available ILI-record, e.g. *kunstproduct* in Dutch (artifact substance) that should be linked to both *artifact* and *product* in English.

HAS_EQ_HYPONYM: a synset can only be linked to more specific ILI-records., *dedo* in Spanish that refers to either a *finger* or *toe* in English.

The EQ_NEAR_SYNONYM relation is typically used for fuzzy concepts and fuzzy matchings. The HAS_EQ_HYPERONYM and HAS_EQ_HYPONYM relations are typically used if there is a gap in English. In addition to these relations there are other less important relation that correspond with each type of language-internal relation: EQ_HAS_HOLONYM, EQ_IN_MANNER, EQ_BE_IN_STATE, EQ_HAS_MERONYM, EQ_CAUSES, EQ_IS_STATE_OF, EQ_INVOLVED, EQ_IS_CAUSED_BY, EQ_ROLE, EQ_HAS_SUBEVENT, EQ_CO_ROLE, EQ_IS_SUBEVENT_OF.

Initially, the ILI was filled with the synsets of the Princeton wordnet (WordNet1.5), but during the project adaptations have been made to make the mapping more efficient. In practice it turns out to be difficult to find a precise matching between a synset in the local wordnet and a synset in the ILI (mostly synsets taken from WordNet1.5). In many cases there will be a many-to-many matching or there will not be an equivalent concept in the ILI. For example, the fuzzy mapping from Dutch to *machine*, *apparatus*, *tool* is also encountered for Italian *utensile* and *ferrovecchio*.

To improve the matching, the ILI is adapted in two ways:

1. Adding of new concepts which are missing

2. Creating sense-groups between closely related senses or regular polysemy

The addition of new concepts is necessary to enable a precise mapping of synsets across wordnets in cases that there is no such concept in WordNet1.5. For example, if only the Spanish and Italian wordnet include a meaning for some type of *wine*, the new concepts should make it possible to specify the equivalence between Spanish and Italian despite the absence in English. First experiments have shown that most of the gaps in the ILI are based on compounds and derivations in the languages that are compositional and productive in meaning, e.g.:

Dutch, doodslaan$_V$	beat to death	totschlagen$_V$, *German*
Dutch, doodstampen$_V$	kick to death	tottrampeln$_V$, *German*
Dutch, casière$_N$	female cashier	cajera$_N$, *Spanish*

Even though these concepts are not directly present in the ILI, it is possible to exhaustively link these concepts to multiple ILI-records. Cases of non-productive gaps occur less frequent but are also more difficult to detect. Especially, finding cases where two languages share the same gap requires rich descriptions and high multilingual skills.

The sense-groups are necessary to deal with inconsistent and fuzzy sense-differentiation across the lexical resources. We often see that resources only specify one out of several meanings that can be distinguished (often on a regular basis): e.g. "embassy" as an *institute* or as a *building* (Buitelaar 1998). This may mean that concepts cannot be linked across languages because different meanings are represented: i.e. either the *institute* or the *building*. To relate these meanings across the wordnets, we extend the ILI with so-called Composite ILI-records that group these meanings: "embasssy", both as a *building* and an *institute*. Each synset in the local wordnet that is linked to one of the more specific meanings will then get an additional equivalence link to a Composite ILI-record (generated automatically). These equivalence relations (EQ_DIATHESIS, EQ_METONYM, EQ_GENERALIZATION) are differentiated from the normal equivalence relations so that it is possible to use these more global matches if a more precise matching gives no result. These adaptations did not break the mapping with WordNet1.5. Added Composite ILI-records:

- **Nouns** 2895 clustered synsets (4,6% of 62780 WN1.5 noun synsets), intersection increased from 7736 (23,8%) to 8183 (25,2%) out of the union of 32520 synsets

- **Verbs** 3839 clustered synsets (31,4% of 12215 WN1.5 verb synsets), intersection increased from 1632 (21,9%) to 3051 (40,9%) out of the union of 7455 synsets

The Composite ILIs have been added to the database independently of all the languages linked to it. This shows how the database can be manipulated and further engineered without having to re-consult all the builders of the wordnets.

3 Top-down building of wordnets.

A drawback of the flexible design described above is that the interpretation and coverage of the wordnets may easily drift apart. There is no guarantee that we cover the same conceptual areas or that we encode the relations in the same way. To minimalize this danger, the wordnets are developed tow-down starting with a shared set of so-called Base Concepts. These Base Concepts have been selected for their importance in the local wordnets. Importance has been measured in terms of the number of relations and the position in the hierarchy. The more relations or the higher the position, the more important a meaning is. All meanings that play a major role in at least two wordnets have been selected. This has resulted in a set of 1059 Base Concepts, represented as WordNet1.5 synsets. The Base Concepts have been described using a top-ontology with 63 basic semantic distinctions (Top-Concepts) such as Substance, Object, Artifact, Natural, Function, Dynamic, Static, Cause, Location, Experience. The top-ontology has been based on other available ontologies and has been adapted to reflect the diversity of the Base Concept selection. The classification of the Base Concepts in terms of the Top-Ontology provides a common framework for the development of the individual wordnets by the different sites.

The actual building of the separate wordnets then takes place along the following steps:

1. The selection of a well-defined set of word meanings.
2. The encoding of lexical semantic relations and equivalence relations for this set.
3. Converting the data to the EuroWordNet import format.
4. Loading the data in the EuroWordNet database.
5. Comparing the wordnets for particular subsets.
6. Revising the wordnets in the EuroWordNet database.
7. Extending the first selection.

First, each group has determined the synsets that most closely represent the common Base Concepts in their local language, given the available resources. This selection has been extended with other meanings which are important in the local wordnets but which are not part of the common set of Base Concepts. This set of meanings has been classified in the local wordnets in terms of their hyperonyms, resulting in a unified tree. Note that these classifications may be different from wordnet to wordnet and still be compatible with the top-ontology classification. In addition to this top-layer, we have included those hyponyms that are also (important) hyperonyms of more specific meanings. Together this selection represents the core of each wordnet with the most important meanings on which the remainder of the vocabulary depends. To summarize, each core wordnet includes at least:

1. The best representatives for the 1059 Base Concepts.
2. Other meanings important for the local wordnet.
3. Hyperonyms for the local Base Concepts.
4. Most important hyponyms of the local Base Concepts.

The core wordnets are specified at least for synonymy, hyponymy and their equivalence relation to the ILI. Optionally, any other salient relation has been encoded to *interconnect* the meanings in the wordnet. Because of the importance for the total wordnets, the manual work has been focused on these cores. The extension from the core wordnets has been done in a top-down direction, using semi-automatic techniques.

The resulting EuroWordNet database is distributed by ELRA (http://www.icp.grenet.fr/ELRA/home.html) on the basis of license agreements. The wordnets require between 10 and 25 MB disk space each. Another 70MB are needed for WordNet1.5 and the Inter-Lingual-Index. All data can also be accessed from CD. 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
Entries	= number of words
LIRels.	= number language-internal relations
EQRels-ILI	= number of equivalence relations to the Inter-Lingual-Index
Synsets without ILI	= synsets without a equivalence relation

Table 2: Quantitative overview of the EuroWordNet database

		Synsets	No. of senses	Entries	LIRels.	EQRels-ILI	Synsets without ILI
Dutch Wordnet	Nouns	34455	54428	45972	84869	26724	6070
	Verbs	9040	14151	8826	25973	26724	1133
	Other	520	1622	1485	797	n.a.	n.a.
Spanish Wordnet	Nouns	18577	41292	23216	40559	18634	0
	Verbs	2602	6795	2278	3749	2602	0
	Other	2191	2439	2439	10855	n.a.	n.a.
Italian Wordnet	Nouns	30169	34552	24903	83021	43848	98
	Verbs	8796	12473	6607	30757	27941	0
	Other	1463	1474	1468	3290	n.a.	n.a.
French Wordnet	Nouns	17826	24499	14879	39172	17815	16
	Verbs	4919	8310	3898	10322	4915	4
	Other	0	n.a.	n.a.	n.a.	n.a.	n.a.
German Wordnet	Nouns	9951	13656	12746	23856	10570	0
	Verbs	5166	6778	4333	10960	5762	0
	Other	15	19	19	2	15	0
Czech Wordnet	Nouns	9727	13829	9277	19856	9729	0
	Verbs	3097	6120	3006	6403	3097	0
	Other	0	n.a.	n.a.	n.a.	n.a.	n.a.
Estonian Wordnet	Nouns	5028	8226	7209	10873	5683	0
	Verbs	2650	5613	3752	5445	3321	0
	Other	0	n.a.	n.a.	n.a.	n.a.	n.a.
English WordNet Addition	Nouns	4751	14188	2524	20707	n.a.	n.a.
	Verbs	11363	25761	14726	21070	n.a.	n.a.
	Other	247	639	70	363	n.a.	n.a.
WordNet1.5, in EuroWordNet format	Nouns	60521	107428	88175	159223	n.a.	n.a.
	Verbs	11363	25768	14734	24331	n.a.	n.a.
	Other	22631	54406	23708	27821	n.a.	n.a.

4 Global Wordnet Association

Most of the groups in EuroWordNet have continued to work on their database. Also other people and groups have taken the initiative to develop a wordnet for their language, often using the same format and structure as EuroWordNet. Furthermore, number of new projects have been launched that continue work within the EuroWordNet model:

- The EUROTERM project extends the EuroWordNet database with specialized terminology (Stamou et al. 2002, http://www.ceid.upatras.gr/Euroterm/).
- The BALKANET project extends the database with more European languages: Czech,
- Romanian, Greek, Turkish, Bulgarian, and Serbian (Stamou et al. 2002, http://www.ceid.upatras.gr/Balkanet/)

- The MEANING project extends the database with sense-tagged corpora extracted from the WWW and word-sense-disambiguation modules (Rigau et al. 2002, http://www.lsi.upc.es/~nlp/projectes/meaning.html).

Although EuroWordNet ended, there is thus still a need to continue a common framework for developing wordnets that can be exchanged and interlinked. Many groups requested help, guidance or information for developing wordnets for their own language. Besides, the discussion on the mapping of vocabulary to concepts is only empirically sound if also non-Western languages are included. Therefore, the coordinators of the Princeton wordnet and EuroWordNet decided to start the Global WordNet Association (http://www.globalwordnet.org) in 2000. The Global WordNet Association (GWA) is a non-profit organization that aims to distribute information and promote communication among researchers that use wordnets, and to co-ordinate efforts in building new wordnets. In particular, the GWA attempts to develop and promote methodologies, standards and common representations for new wordnets that would allow these resources to connect and communicate. Currently, GWA registered over 30 wordnets that have either been built and are available or are under construction.

References

Buitelaar, P. 1998, Corelex: systemeatic polysemy and underspecification. PhD. Thesis, Department of Computer Science, Brandeis University

Copeland, C., Durand, J., Krauwer, S. and Maegaard, B. (eds.). 1991. *The Eurotra Formal Specifications*, Office for Official Pblications of the European Community, Luxembourg.

Fellbaum C. editor, 1998, *WordNet An Electronic Lexical Database*. The MIT Press.

Magnini B. & Cavaglià G., *Integrating subject field codes into WordNet*. In Proceedings of the Second Internatgional Conference on Language Resources and Evaluation LREC'2000, Athens. Greece. 2000.

Nirenburg, S. (ed.). 1989. "Knowledge-based MT", *Special issue Machine Translation* vol.4, no 1 and 2, Kluwer Publishers, Dordrecht.

Rigau, German., Bernardo. Magnini, Eneko. Agirre, Piek. Vossen and John. Carroll (2002), MEANING: a Roadmap to Knowledge Technologies, COLING 2002, Taipe.

Stamou Sofia, Oflazer Kemal, Pala Karel, Christoudoulakis Dimitris, Cristea Dan, Tufis Dan, Koeva Svetla, Totkov George, Dutoit Dominique, Grigoriadou Maria. 2002a, Balkanet: A Multilingual Semantic Network for the Balkan Languages.

Stamou Sofia, Ntoulas Alexandros, Hoppenbrouwers Jeroen, Saiz-Noeda Maximiliano, Christodoulakis Dimitris. 2002b, Extending the EuroWordNet with Domain- Specific Terminology Using an Expand Model Approach.

Vossen P. editor 1998, *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*, Kluwer Academic Publishers, Dordrecht.

Vossen P., 1999, *EuroWordNet General Document*,.http://www.globalwordnet.org/ewn/general_document.ps