

Definition of the links and subsets for nouns of the EuroWordNet project

Version 6, November, 1996
Final

Salvador Climent
FUE-U. de Barcelona

Horacio Rodríguez
FUE- U. Politècnica de Catalunya

Julio Gonzalo
FUE-UNED



Deliverable D005, WP3.1, EuroWordNet, LE2-4003

* The document compiles work from AMS group related to Dutch, from PSA group related to Italian and from FUE group related to Spanish. In addition, Piek Vossen and Felisa Verdejo have made significant contributions to the contents of this deliverable.

Identification number	LE-4003-D-005
Type	Document
Title	Definition of the links and subsets for nouns
Status	Final
Deliverable	D005
Work Package	WP3
Task	T3.1
Period covered	June-August1996
Date	07/11/96
Version	6
Number of pages	123
Authors	Salvador Climent Horacio Rodríguez Julio Gonzalo
WP/Task responsible	FUE
Project contact point	Piek Vossen Computer Centrum Letteren University of Amsterdam Spuistraat 134 1012 VB Amsterdam The Netherlands tel. +31 20 525 4624 fax. +31 20 525 4429 e-mail: Piek.Vossen@let.uva.nl
EC project officer	Jose Soler
Status	Public
Actual distribution	Project Consortium
Supplementary notes	n.a.

Key words	Semantic databases, Information Retrieval, Language Engineering, Semantic Relations, Semantic Test Sentences
Abstract	This document sets up the basis to build individual wordnets for nouns in every language included in the EuroWordNet project: Dutch, Italian and Spanish. Two main topics are covered: one is the definition of all possible relations between nouns; the other is the construction of initial subsets for every language. The document also reports the experiments carried out on specific semantic domains to foresee the main issues and problems when building the whole EuroWordNet database for nouns.
Status of the abstract	Complete
Received on	
Recipient's catalogue number	

Executive Summary

In this document we give a description of the semantic relations for nouns and a definition of the covered vocabulary. The purpose of the specification of the relations and the covered vocabulary is to provide a common-framework for building the multilingual wordnets at the different sites. The results of an experiment - consisting of building microwordnets for some concrete semantic domains with our sources and tools - are also reported.

The relations

The semantic relations are based on the relations in WordNet1.5 with some additions and changes motivated in D001. This document covers the language-internal relations. Bi-lingual relations - except for a *translation* relation - are in principle the same as the monolingual relations, whereas the language-external relations (relation with and between domains, relations with and between top-concepts and relations with and between instances) will be defined at a later stage. Another issue which is not covered in this document is the way these relations will be extracted. This will be dealt with in WP2, D021 and D025 (The EuroWordNet Tools Report).

The monolingual relations are defined in terms of:

- Non ambiguous descriptions covering all relations of interest, that serve as a common framework for all sites from the beginning of the construction process.
- Linguistic tests for every relation, in every language, that permit fast and reliable verification of the database quality and consistency. That tests consist of:
 1. semantic substitution contexts for word pairs.
 2. specific conditions on the word pairs.
 3. examples.
 4. a specification of the effect in terms of an explicit EuroWordNet relation.
- Sets of constraints for every relation that facilitate automatic consistency checking within the database.

The list is to be interpreted as an exhaustive description of all phenomena of interest to the EWN database, not as a list of every relation that must be completely covered in the database for every language. The importance of each relation in the database will depend on the sources of each site and their ability to extract the adequate information from such sources.

The Subsets

The vocabulary covered in EuroWordNet is first described in terms of general criteria. In order to achieve maximal overlap between the three languages, the building of the monolingual databases is split into two sub-tasks. The first one is a manual construction of the relations for the base concepts. The second one is the automatic extension with more specific concepts linked to that base concepts.

This section concentrates on describing the criteria to select the base concepts and the main guidelines to build the subset of base concepts for each language. This methodology involves the exchange of preliminary base concepts between the sites. A comparison and evaluation of these

first selections have been carried out to establish how to modify/extend each monolingual subset to maximize the overlap with the others. The state of completion of this task is reported.

The Experiments

The third section describes the construction of wordnets for some small fragments. This experiment was set up to verify the initial set of relations, to anticipate problems and discuss solutions. Nominal wordnet-fragments have been built for the semantic fields “food”, “drink”, “feeling” and “instrument”. The experiments show that fully automatic extraction methods are not still reliable, and lead to mismatches between synonymy and hyperonymy/hyponymy. However, this approach, combined with some manual selection of the appropriate intermediate nodes on automatically derived taxonomies, leads to good results. As most of the problems come from the more abstract - and general - concepts, we expect that the manual construction of the base concepts will make automatic strategies more reliable when attaching more specific concepts.

Annexes

The annexes give detailed information of the work carried out in each site. They include a report on the extraction of base concepts for Dutch and Italian, the intersection list of base concepts for Dutch and Spanish, a detailed description of the individual experiments in each site, and the results of an experiment on acquiring interlingual synonymy between Spanish and English.

Table of Contents

1. Introduction and motivation.....	1
2. Definition of relations	3
2.1 Synonymy and the notion of Synset	9
2.1.1 Test for synonymy (synset membership)	10
2.1.2 NEAR_SYNONYM.....	11
2.2 Hyperonymy and Hyponymy	19
2.2.1 HAS_HYPONYM and HAS_HYPERONYM	20
2.2.2 HAS_XPOS_HYPERONYM and HAS_XPOS_HYPONYM	23
2.3 Antonymy	32
2.4 Part-whole Relations	34
2.4.1 HAS_MERONYM and HAS_HOLONYM	35
2.4.2 HAS_HOLO_MEMBER and HAS_MERO_MEMBER	37
2.4.3 HAS_HOLO_PORTION and HAS_MERO_PORTION.....	39
2.4.4 HAS_HOLO_PART and HAS_MERO_PART.....	41
2.4.5 HAS_HOLO_MADEOF and HAS_MERO_MADEOF	44
2.4.6 HAS_HOLO_LOCATION and HAS_MERO_LOCATION	45
2.5 “Role” and “Involved” relations and their subtypes.....	46
2.6 BE_IN_STATE & STATE_OF.....	49
2.7 Instance / Class Relations	50
2.8 Undefined Relations: “fuzzynoms”	51
2.9 Interlingual Relations.....	51
3. Definition of the initial subset of base concepts	52
4. Determining the subset	55
5. Base-concepts exchange	57
6. Experiments	62
6.1 Introduction	62
6.2 Major conclusions	63
6.3 Result of experiment for Dutch.....	64
6.4 Result of experiment for Italian.....	65
6.5 Result of experiment for Spanish	68
Appendix I: Extraction of base concepts	70
I.1 Extraction of Dutch subset.....	70
I.2 Extraction of Italian subset.....	76
I.3 Intersection between Spanish and Dutch subset.....	79
Appendix II: Experiments on specific subsets.....	82
II.1 Detailed description of Dutch experiment.....	82
II.1.1. Introduction	82
II.1.1.1 The sense-distinctions.....	83
II.1.1.2 Synonymy	84
II.1.1.3 Hyperonymy.....	86
II.1.1.4. Hyponymy.....	87

II.1.1.5 Antonymy	89
II.1.1.6 Associative.....	90
II.1.1.7 Reference	91
II.1.1.8 Partitive.....	91
II.1.1.9 Woman	92
II.1.1.11 Preference	93
II.1.1.12 Symbol.....	93
II.1.1.13 Abbreviation.....	93
II.1.1.14 Form-variant	94
II.1.2. Extracting information.....	94
II.2 Detailed description of Italian experiment	102
II.2.1 Scanning the data	102
II.2.2 Frequency lists.....	104
II.2.3 Structuring the taxonomies	106
II.2.4. Reorganization of taxonomies in synsets	108
II.2.5 Other semantic relations	111
II.3 Detailed description of Spanish experiment	114
II.3.1 Estimation of the taxonomy as source of relations.....	114
II.3.2 Results of the experiments	115
Appendix III Experiments on discovering interlingual synonymy.....	120
REFERENCES	123

1. Introduction and motivation

This document sets up the basis to build individual wordnets for nouns in every language covered by the EuroWordNet (EWN) project: Dutch, Italian and Spanish. Two main topics are covered: one is the definition of all possible relations between nouns, the other is the construction of initial subsets for every language.

The description of relations between nouns in this document does not provide formal definitions. However, it is intended to provide:

- **non ambiguous descriptions** covering all relations of interest, that serve as a common framework for all sites from the beginning of the construction process.
- **linguistic tests** for every relation, in every language, that permit fast and reliable verification of the database quality and consistency.
- **sets of constraints** for every relation that facilitate automatic consistency checkings within the database.

The list is to be interpreted as an exhaustive description of all phenomena of interest to the EWN database, not as a list of every relation that must be completely covered in the database for every language; the importance of each relation in the database will depend on the sources of each site and their ability to extract the adequate information from such sources. Special attention is paid to clarify the concepts of synonymy and hyponymy-hyperonymy in the EWN database, as they are the key concepts that give a structure to the database.

Besides enumerating and describing relations between nouns, an experiment has been carried out to test each site's tools and resources to build the wordnets. The experiments consist of building microwordnets for some specific semantic domains (the same domains for all partners). These experiments permit to estimate future results and foresee major problems while constructing the wordnets from our sources.

The construction of initial subsets is also important to set up a common framework to build the individual wordnets. The aim is to create a core subset for each language that contains the most relevant synsets (according to the number of relations per synset and their position in the hierarchy) and is highly overlapped with the others. Manual construction efforts in EWN will concentrate on getting high quality, generic initial subsets with a balanced, multilingual coverage. These subsets will serve as an anchor for the rest of the synsets (usually formed by more specific, less polysemic words that are easier to treat automatically). That methodology will hopefully guarantee that the resulting wordnets for each language will be maximally balanced and compatible, facilitating a true multilingual usage of the database. This document describes a common methodology to build that initial subsets and to use them as a seed for the whole noun database. The state of completion of this methodology is also reported.

The selections, relations and relation tests described in this deliverable are, within reasonable limits, subject to some change during the project. Certainly, new data and phenomena will come out when developing the WordNets, and some modifications will be necessary. The results coming out of this deliverable will be put in the WWW-site of the project

(<http://www.let.uva.nl/CCL/EuroWordNet.html>). In particular, for every language there will be available:

- 1) first list of base concepts
- 2) first list of relations
- 3) first definitions of relations with tests

These results will be subject to updates.

This deliverable is organized as follows:

- Next section is devoted to the definition of the available relations in the EWN database.
- Section 3 establishes the conditions that the initial subsets must observe.
- Section 4 discusses the methodology to build the EWN noun database.
- Section 5 discusses the result of the exchange of initial concepts between the partners, which is a crucial step to maximize overlapping between them.
- Section 6 evaluates the experiments on particular semantic domains to anticipate the major problems that will arise when building the whole wordnets.
- Finally, some appendices give details on the experiences and results of every site. Appendix 1 covers the selection of subsets and the intersection between Dutch and Spanish subsets. Appendix 2 reports on the experiments covering specific semantic domains. Finally, Appendix 3 describes an experiment of FUE to evaluate the possibility of automatically establishing interlingual synonymy links.

2. Definition of relations

This section is devoted to all possible relations between nouns in the EuroWordnet database, and some relations from nouns to verbs and nouns to adjectives and adverbs. In particular, we consider language-internal relations, and we assume that interlingual relations (between each language and wordnet 1.5, which acts as an interlingual connection) are equivalent. For instance, `HAS_HYPERONYM` and `HAS_EQ_HYPERONYM` can be defined in the same way, with the sole exception that the second one holds among synsets of different languages.

A closely related problem is how to extract in a semi-automatic way this information from sources such as Machine Readable Dictionaries (MRDs). One way to proceed would consist of defining for each relation a set of appropriate dictionary definition patterns from which to retrieve pairs of words standing in such a relation. With this approach, the structure of dictionary definitions would serve as a basis for both defining and testing relations. This subject will be dealt with in detail in next tasks of Work Packages (WP) 2, 3 and 4 of the project.

List of relations

After a period of discussion, the partners have reached an agreement on the relations to be considered. The list below provides, for each relation, its name, the part of speech of the synsets involved, labels applying to the relation and the database types between which the relation may hold. For example, `HAS_XPOS_HYPERONYM` is a relation between word meanings or synsets (WM) with six different combinations of parts of speech, where the first one corresponds to relations between noun and verb synsets, possibly with the labels disjunction or conjunction (Dis/Con) and/or the label reversed. The labels in the third column will be further explained in this document.

Relation Type	Parts of Speech	Labels	Data Types
NEAR_SYNONYM	N<>N, V<>V	reversed	WM<>WM
XPOS_NEAR_SYNONYM	N<>V, N<>AdjAdv, V<>AdjAdv	reversed	WM<>WM
HAS_HYPERONYM	N>N, V>V	Dis, Con, reversed	WM<>WM
HAS_HYPONYM	N>N, V>V	Dis, reversed	WM<>WM
HAS_XPOS_HYPERONYM	N>V, N>AdjAdv, V>AdjAdv, V>N, AdjAdv>N, AdjAdv>V	Dis, Con, reversed	WM<>WM
HAS_XPOS_HYPONYM	N>V, N>AdjAdv, V>AdjAdv, V>N, AdjAdv>N, AdjAdv>V	Dis, reversed	WM<>WM
HAS_HOLONYM	N>N	Dis, Con, reversed, not	WM<>WM
HAS_HOLO_PART	N>N	Dis, Con, reversed, not	WM<>WM
HAS_HOLO_MEMBER	N>N	Dis, Con, reversed, not	WM<>WM
HAS_HOLO_PORTION	N>N	Dis, Con, reversed, not	WM<>WM
HAS_HOLO_MADEOF	N>N	Dis, Con, reversed, not	WM<>WM
HAS_HOLO_LOCATION	N>N	Dis, Con, reversed, not	WM<>WM
HAS_MERONYM	N>N	Dis, Con, reversed, not	WM<>WM
HAS_MERO_PART	N>N	Dis, Con, reversed, not	WM<>WM
HAS_MERO_MEMBER	N>N	Dis, Con, reversed, not	WM<>WM
HAS_MERO_MADEOF	N>N	Dis, Con, reversed, not	WM<>WM
HAS_MERO_LOCATION	N>N	Dis, Con, reversed, not	WM<>WM
ANTONYM	N<>N, V<>V	reversed	VA<>VA
NEAR_ANTONYM	N<>N, V<>V	reversed	WM<>WM
XPOS_NEAR_ANTONYM	N<>V, N<>AdjAdv, V<>AdjAdv	reversed	WM<>WM
CAUSES	V>V, N>V, N>N, V>N, V>AdjAdv, N>AdjAdv	Dis, Con, Factive, reversed, not	WM<>WM
IS_CAUSED_BY	V>V, N>V, N>N, V>N, AdjAdv>V, AdjAdv>N	Dis, Con, Factive, reversed, not	WM<>WM
HAS_SUBEVENT	V>V, N>V, N>N, V>N	Dis, Con, reversed, not	WM<>WM
IS_SUBEVENT_OF	V>V, N>V, N>N, V>N	Dis, Con, reversed, not	WM<>WM
ROLE	N>V, N>N	Dis, Con, reversed	WM<>WM

ROLE_AGENT	N>V, N>N	Dis, Con, reversed	WM<>WM
ROLE_INSTRUMENT	N>V, N>N	Dis, Con, reversed	WM<>WM
ROLE_PATIENT	N>V, N>N	Dis, Con, reversed	WM<>WM
ROLE_LOCATION	N>V, N>N, AdjAdv>N, AdjAdv>V	Dis, Con, reversed	WM<>WM
ROLE_DIRECTION	N>V, N>N, AdjAdv>N, AdjAdv>V	Dis, Con, reversed	WM<>WM
ROLE_SOURCE_DIRECTION	N>V, N>N, AdjAdv>N, AdjAdv>V	Dis, Con, reversed	WM<>WM
ROLE_TARGET_DIRECTION	N>V, N>N, AdjAdv>N, AdjAdv>V	Dis, Con, reversed	WM<>WM
INVOLVED	V>N, N>N	Dis, Con, reversed	WM<>WM
INVOLVED_AGENT	V>N, N>N	Dis, Con, reversed	WM<>WM
INVOLVED_PATIENT	V>N, N>N	Dis, Con, reversed	WM<>WM
INVOLVED_INSTRUMENT	V>N, N>N	Dis, Con, reversed	WM<>WM
INVOLVED_LOCATION	V>N, N>N, N>AdjAdv, V>AdjAdv	Dis, Con, reversed	WM<>WM
INVOLVED_DIRECTION	V>N, N>N, N>AdjAdv, V>AdjAdv	Dis, Con, reversed	WM<>WM
INVOLVED_SOURCE_DIRECTION	V>N, N>N, N>AdjAdv, V>AdjAdv	Dis, Con, reversed	WM<>WM
INVOLVED_TARGET_DIRECTION	V>N, N>N, N>AdjAdv, V>AdjAdv	Dis, Con, reversed	WM<>WM
BE_IN_STATE	N>AdjAdv, V>AdjAdv	Dis, Con, reversed, not	WM<>WM
STATE_OF	AdjAdv>N, AdjAdv>V	Dis, Con, reversed, not	WM<>WM
FUZZYNYM	N<>N, V<>V		WM<>WM
XPOS_FUZZYNYM	N<>V, N<>AdjAdv, V<>AdjAdv		WM<>WM
EQ_SYNONYM	N<>N, V<>V		WM<>ILIR
EQ_NEAR_SYNONYM	N<>N, V<>V, N<>AdjAdv, V<>AdjAdv,		WM<>ILIR
			WM<>ILIR
HAS_EQ_HYPERONYM	N>N, N>V, N>AdjAdv, V>V, V>N, V>AdjAdv, AdjAdv>N, AdjAdv>V		WM<>ILIR
HAS_EQ_HYPONYM	N>N, N>V, N>AdjAdv, V>V, V>N, V>AdjAdv, AdjAdv>N, AdjAdv>V		WM<>ILIR
HAS_EQ_HOLONYM	N>N		WM<>ILIR
HAS_EQ_MERONYM	N>N		WM<>ILIR
			WM<>ILIR
EQ_INVOLVED	N>N, V>N		WM<>ILIR
EQ_ROLE	N>N, N>V		WM<>ILIR
			WM<>ILIR
EQ_CAUSES	N>N, V>V, N>V, V>N, V>AdjAdv, N>AdjAdv		WM<>ILIR
EQ_IS_CAUSED_BY	N>N, V>V, N>V, V>N, AdjAdv>V, AdjAdvV>N		WM<>ILIR
EQ_HAS_SUBEVENT	N>N, V>V, N>V, V>N		WM<>ILIR

EQ_IS_SUBEVENT_OF	N>N, V>V, N>V, V>N		WM<>ILIR
EQ_BE_IN_STATE	V>AdjAdv, N>AdjAdv		WM<>ILIR
EQ_IS_STATE_OF	AdjAdv>V, AdjAdv>N,		WM<>ILIR
HAS_INSTANCE	N>PN		WM>I
BELONGS_TO_CLASS	PN>N		I>WM

Parts of Speech:

N	= noun
V	= verb
AdjAdv	= Adjective or Adverb
PN	= pronoun or name

Labels:

factive/non-factive	= restricted to Lyon's distinction between factive causes and modally more complex causes such as <i>search-find</i>								
reversed	= for relations which are not explicitly stated but generated by automatic reversal, e.g. <i>kill</i> causes <i>die</i> whereas <i>die</i> is caused by <i>kill</i> only as the reversal of this relation.								
Dis/Con	= for explicitly coded disjunction/conjunction of relations, e.g. an <i>airplane</i> has <i>jets</i> or <i>propellers</i> (disjuncts) and <i>wings</i> and a <i>tail</i> (conjuncts).								
not	= for cases in which the negative implication holds between word pairs, e.g.: <table> <tr> <td>synset</td> <td>relation</td> <td>label</td> <td>synset</td> </tr> <tr> <td>{ape}</td> <td>HAS_MERONYM:</td> <td>not</td> <td>{tail}</td> </tr> </table>	synset	relation	label	synset	{ape}	HAS_MERONYM:	not	{tail}
synset	relation	label	synset						
{ape}	HAS_MERONYM:	not	{tail}						

Data types:

WM	= word meaning or synset
VA	= variant or synset member
I	= instance
ILIR	= ILI record

List of Interlingual relations

The same list as above, but the relations apply always between Dutch/Italian/Spanish synsets and WN1.5 synsets. Only one specific interlingual relation is considered:

TRANSLATES_AS (N<>N, V<>V, N<>V, N<>A, V<>A) [Word Forms]

This relation expresses a direct translation relation between specific word senses or variants within a synset. As such, it does not hold between synsets (which is the normal equivalence relation) but between specific forms or variants within the synsets. Since this information cannot be accessed by the system as the other relations, it will not be stored as an operational semantic relation between synsets but as an information field for each variant (see D007 "The Architecture of the EuroWordNet database"). Each variant in a synset may thus have one or more translations in one or more languages.

Definitions and tests: general approach

First of all, it must be noticed that EWN is a linguistic database, not a world-knowledge database, although such structures are not necessarily incompatible -in fact the linguistic database could be used as a basis for eventually deriving a world-knowledge database. A linguistic database implies that relations are encoded between words, or incidentally idioms, as lexicalized units in a language system. From a knowledge engineering point of view it may very well be that particular levels are needed or efficient but these may not be lexicalized in the vocabulary of a language; for instance, "finger" and "toe" may be seen as separate body parts in a knowledge base, but Spanish only lexicalizes "dedo", a word which covers both knowledge-base-objects. In that case the Spanish wordnet will only include "dedo" -although its relation to both English concepts "finger" and "toe" may be inferred via interlingual relationships.

The EuroWordNet database can to some extent be used for making inferences. However, the implications expressed by the relations should not be seen as logical consequences at all costs but as typically expected consequences from a common-sense perspective. In this respect the implicational consequence also varies from relation type to relation type. The HAS_HYPERONYM relation (e.g. "it is a dog therefore it is an animal") will often be more strict than e.g. the ROLE_INSTRUMENT relation (e.g. "it is a hammer therefore is used for hammering"). Making inferences on the basis of the coded relations should therefore be done with caution. Nevertheless, the database is still very useful for less delicate tasks like information retrieval and it can still be used as a basis for the construction of a database in which the implications are differentiated in terms of their consequence and the inferences are more clear.

It is a requirement of the database that every relation has a reverse counter-part. However, there is a difference between relations which are explicitly coded as reverse relations and relations which are automatically reversed because of this requirement. For example, it may very well be the case that "finger" is defined by reference to "hand" and "hand" as a body part consisting of fingers. In that case the relation is also conceptually bi-directional . However, if we say that a "paper-clip" is made of "metal" then the reverse that "metal" can be shaped into a "paper-clip" seems to be of another order. This becomes even more clear for parts which can also occur independently of some whole, e.g. "time bomb" which may be linked to the meronym "clock" but the reversal to say that a "clock" is part of a "time-bomb" is less salient. To be able to distinguish between the conceptually-dependent and the automatically-reversed relations we use the label **reversed**.

Furthermore, there is a possibility to have a negation label **not** to explicitly express that a relation does not hold. Such a negative relation is not necessarily the same as antonymy, which expresses opposition. This becomes clear when we consider negative meronymy (and sub-event relations), where a holonym is said to have not a part. A classical example in English is in this respect the difference between "ape" and "monkey" where the latter refers to these animals in general (with or without a tail) and the former to the subclass without a tail. This can be explicitly expressed using the negation label as follows, whereas we cannot express this as antonymy:

synset	relation	label	synset
{ape}	HAS_MERONYM:	not	{tail}

Finally, as explained in D001 relations may form disjunctive or conjunctive sets. In the case of meronymy this is particularly salient when a holonym consists of multiple meronyms making up the

whole together (conjoined by default) or when particular meronyms such as e.g. “wheel” can occur in alternative holonyms such as “car” and “bicycle” (explicitly disjunctive choices).

Similarly, as pointed out above EWN will not encode all possible properties and features which can be involved in a detailed description of a lexical concept but only the lexical semantic relations which, after discussions within the group, have turned out to be both the most basic from a lexical-semantic point of view and, as well, more easily retrievable from the participant's sources.

A method for the verification of lexical-semantic relations is needed. The most consistent way of doing this, according to linguistic literature, is using substitution tests or diagnostic frames. Substitution tests make the relation between words explicit arousing a clear semantic intuition comparable with grammaticality intuitions of sentences. For example, Cruse (1986) uses the following frame to distinguish **hyponymy** and **synonymy** relations between words:

Synonyms are mutually substitutable in the following sentence:

It is a fiddle therefore it is a violin
It is violin therefore it is a fiddle

Hyponyms and *hyperonyms* are only partially substitutable:

It is a dog therefore it is an animal
*It is animal therefore it is a dog

Similar sentences can be formulated for other relations. The substitution context is semantic and is not intended to cover syntactic differences, differences in register, style or dialect nor pragmatic differences between words. Therefore, the major criterion for setting relations in EWN is the *Substitution principle* -or a comparable linguistic test, in the case the former is not suitable; definition patterns, for instance, can also be seen as test sentences or diagnostic tests.

In addition, a principle of economy will also be taken into account:

If a word W1 is the hyperonym of W2 and W2 is the hyperonym of W3 then W3 should not be linked to W1 but to W2.

This principle should prevent that senses are linked too high up the hierarchy and intermediate levels are skipped. It can also be used for other relations such as meronymy and cause.

A third restriction is the principle of homogeneity: two words cannot be related via different relations types:

If a word W1 is related to W2 via relation R1, W1 and W2 cannot be related via relation Rn, where Rn is defined as a distinct relation from R1.

By definition, all relations are incompatible, except for those relations which are defined as subtypes of a more general relation (see below). The exceptions are *hyperonymy-hyponymy* and *set-membership* since in some cases the later may be seen as a sort of class-inclusion -which is the logical relation encoded by the former. Therefore, HAS_HYPERONYM and

HAS_HOLO_MEMBER on the one side and HAS_HYPONYM and HAS_MERO_MEMBER on the other, are not incompatible.

These tests do not always result absolute values: yes or no. Therefore we allow for the following scores: yes, probably, unclear, unlikely, no. This means that in problematic cases a word can still be related but with less explicit links. The scalar values can also be used to differentiate manual decisions and explicit automatic linking from less reliable default relations.

In the next sections, all possible relations between nouns and relations from nouns to verbs and noun to adjectives/adverbs are described. For each relation, we will give test-sentences that can be applied to any pair of words to verify the coded relations. The tests are expressed first in English, and then in the other languages included in the project: Dutch, Italian and Spanish.

2.1 Synonymy and the notion of Synset

Miller and Fellbaum (1992) report that using synonymy definition attributed to Leibniz (*two expressions are synonyms if the substitution of one for the other never change the truth value of a sentence in which the substitution is made*) true synonyms would be rarely found. Therefore, they suggest to make use of a weak notion of synonymy, namely '**semantic similarity**', which would entail interchangeability of two words in a given context. Furthermore, Cruse (1986) reports that 'there is, unfortunately, no neat way of characterising synonyms'. Moreover, in some cases, the intuitive notions of synonymy and hypernymy-hyponymy (namely class inclusion) may be confused. Following Miller and Fellbaum (1992), what seems clear is that synonymy should be a symmetric relation, that is, if X is 'semantically similar' to Y, then Y is equally 'semantically similar' to X, while, obviously, hypernymy-hyponymy should be asymmetric.

Synonymy is the basis for the organization of the database in synsets. In principle all semantically equivalent words should belong to the same synsets (where they can be differentiated by labels on the appropriate usage). By **semantically-equivalent** we mean that two words denote the same range of entities, irrespective of the morpho-syntactic differences, differences in register, style or dialect or differences in pragmatic use of the words. Another, more practical, criterion which follows from the homogeneity principle is that two words which are synonymous cannot be related by any of the other semantic relations defined.

This would mean that, for example, the following variants belong to the same synset:

{people, folks}
{cop, pig, policeman, police officer}

but it also means that a pair such as "jewel" and "jewelry" cannot belong to the same synset because there is another semantic relation: "**member-group**" that can be used to relate them (even though they are to a large extent interchangeable in language use).

Strictly speaking, this definition allows for synonymy across parts-of-speech, e.g. "shot N", "shoot V". However, since the distinction between part-of-speech (as an intrinsic property of WordNet1.5) is crucial to many systems using WordNet1.5 we have decided to use a separate relation for synonymy (and also hyponymy) across parts-of-speech: XPOS_NEAR_SYNONYM.

Except for the above cases variants in the synset should be fully substitutable in a semantic context. This means in general that, for nouns:

- in any sentence *S* where Noun1 is the head of an NP which is used to identify an entity in discourse another noun Noun2 which is a synonym of Noun1 can be used as the head of the same NP without resulting in semantic anomaly.
- in any sentence *S* where Noun2 is the head of an NP which is used to identify an entity in discourse another noun Noun1 which is a synonym of Noun2 can be used as the head of the same NP without resulting in semantic anomaly.

The substitution sentences for synonymy are the same as for hyponymy. They differ in the fact that synonyms are mutually exclusive whereas words with a hyponymy relation are partially interchangeable (see below). Next the specific test-sentences for the synonymy relations are given. The tests are numbered and each numbered test can consist of multiple sentences to be applied simultaneously. Furthermore, in some cases an additional condition is formulated. In addition to the generic test sentence an example is given and the effect is specified.

2.1.1 Test for synonymy (synset membership)

Test 1

<i>Comment:</i>		Synonymy of nouns
<i>Score</i>		<i>Test sentence</i>
yes	a	if it is (a/an) X then it is also (a/an) Y
yes	b	if it is (a/an) Y then it is also (a/an) X
<i>Conditions:</i>		X and Y are singular or plural nouns
<i>Example:</i>	a	if it is a fiddle then it is a violin
	b	if it is a violin then it is a fiddle
<i>Effect:</i>		synset variants {fiddle, violin}

Test 1-Dutch

<i>Score</i>		<i>Test sentence</i>
yes	a	als het een X is dan is het ook een Y
yes	b	als het een Y is dan is het ook een X
<i>Conditions:</i>		X and Y are singular or plural nouns
<i>Example:</i>	a	als het een fiets is dan is het ook een rijwiel
	b	als het een rijwiel is dan is het ook een fiets
		X = fiets (bike)
		Y = rijwiel (bicycle)
<i>Effect:</i>		synset variants {fiets, rijwiel}

Test 1-Spanish

<i>Score</i>		<i>Test sentence</i>
yes	a	Es (un) X, luego es (un) Y
yes	b	Es (un) Y, luego es (un) X.
<i>Conditions:</i>		X and Y are singular nouns

Example: a Es un mechero luego es un encendedor
 b Es un encendedor luego es un mechero
 X = mechero (lighter)
 Y = encendedor (lighter)

Effect: synset variants {encendedor, mechero}

Test 1-Italian

Score Test sentence

yes a É(un, una) X, allora é(un, una) Y
 yes b É(un, una) allora é(un, una) X

Conditions: X and Y are singular or plural nouns

Example: a É una dalia allora é una giorgina
 b É una giorgina allora é una dalia
 X = dalia (dahlia)
 Y = giorgina (dahlia)

Effect: synset variants {dalia, giorgina}

2.1.2 NEAR_SYNONYM

From the experiments it followed that in many cases there is a close relation between words but not sufficient to make them members of the same synset. That they are not genuine synonyms follows from the fact that they do not yield extreme scores for the previous test or that the hyponyms linked to each of these word cannot be exchanged (see the examples in the experiments). Especially in the latter case we would like to keep the hyponyms separate and still be able to express that the two words are close in meaning. For this purpose we can make use of the NEAR_SYNONYM relation. Near-synonyms can be detected either by using Test 1 and allowing not only *yes* scores but also *probably* (Test 2) or by using a test (Test 3 below) to elicit differences in denotational range of more or less equivalent words.

Test 2

Comment: Near-Synonymy of nouns

Score Test sentence

yes/probably a if it is (a/an) X then it is also (a/an) Y
 yes/probably b if it is (a/an) Y then it is also (a/an) X

Conditions: at least one of the scores must be *probably*

Example: a if it is garbage then it is trash
 b if it is trash then it is garbage

Effect: garbage NEAR_SYNONYM trash
 trash NEAR_SYNONYM garbage

Test 2- Dutch

Score Test sentence

yes/probably a als het een X is dan is het ook een Y
 yes/probably b als het een Y is dan is het ook een X

Conditions: X and Y are singular or plural nouns

Example: a als het een gevoel is dan is het ook een stemming
 b als het een stemming is dan is het ook een gevoel

Effect: X =gevoel (feeling)
 Y =stemming (mood)
 gevoel NEAR_SYNONYM stemming
 stemming NEAR_SYNONYM gevoel

Test 2-Spanish

Score Test sentence
 yes/probably a Es un X, luego es un Y
 yes/probably b Es un Y, luego es un X.
 Conditions: at least one of the scores must be *probably*
 Example:
 yes a Es un aguardiente luego es un licor
 probably b Es un licor luego es un aguardiente
 Effect: licor (liquor) NEAR_SYNONYM aguardiente (spirits)
 aguardiente (spirits) NEAR_SYNONYM licor (liquor)

Test 2-Italian

Score Test sentence
 yes/probably a se (é/sono) X allora (é/sono) anche Y
 yes/probably b se (é/sono) Y allora (é/sono) anche X
 Conditions: at least one of the scores must be *probably*
 Example: a se sono immondizie allora sono rifiuti
 b se sono rifiuti probabilmente sono immondizie
 X = immondizie (dirt)
 Y = rifiuti (rubbish)
 Effect: rifiuti NEAR_SYNONYM immondizie
 immondizie NEAR_SYNONYM rifiuti

Test 3

Comment: Measures differences in denotational range of close concepts.

Score *Test sentence*

yes a if it is a/an X then it is also a/an Y but you say X only of Z_n

yes b if it is a/an Y then it is also a/an X but you say X only of Z_n

Conditions: Z_n are hyponyms of X

Example: a if it is a tool then it is also an instrument but you say it only of hammers, screw drivers, etc.

b (reverse)

Effect: tools NEAR_SYNONYM instrument

instrument NEAR_SYNONYM tools

Test 3-Dutch

Score *Test sentence*

yes a als het een X is dan is het ook een Y maar je zegt het alleen van Z_n

yes b als het een Y is dan is het ook een X maar je zegt het alleen van Z_n

Conditions: Z_n are hyponyms of X

Example: a als het (een stuk) gereedschap is dan is het ook een hulpmiddel maar je zegt het alleen van schoevedraaiers, hamers, tangen, etc..

X = gereedschap (tools)

Y = hulpmiddel (means)

Z = schroevendraaiers (screw drivers), hamers (hammers), tangen (tongs), etc..

b als het een hulpmiddel is dan is het ook (een stuk) gereedschap maar je zegt het alleen van niet-electrische apparaten, en voorwerpen.

X = gereedschap (tools)

Y = hulpmiddel (means)

Z = niet-electrische apparaten (non-electric apparatus), voorwerpen (objects)

Effect: gereedschap NEAR_SYNONYM hulpmiddel

hulpmiddel NEAR_SYNONYM gereedschap

Test 3

Score *Test sentence*

yes a si es un X entonces también es un Y, pero esto sólo vale para Z_n

yes b si es un Y entonces también es un X, pero esto sólo vale para Z_n

Conditions: Z_n are hyponyms of X

Example: a si es un instrumento, entonces tambien es una herramienta, pero esto sólo vale para martillos, destornilladores, llaves inglesas, etc.

b (reverse)

X = instrumento (instrument)

Y = herramienta (tool)

Z = instrumentos manuales (manual tools)

Effect: instrumento NEAR_SYNONYM herramienta

herramienta NEAR_SYNONYM instrumento

Test 3-Italian

Score *Test sentence*

yes	a	se é un X allora é anche un Y ma questo vale solo per Z_n
yes	b	se é un Y allora é anche un X ma questo vale solo per Z_n
Conditions:		Z_n are hyponyms of X
Example:	a	se é uno strumento allora é anche un arnese ma questo vale solo per vanghe, zappe, martelli, cacciaviti, forbici, pinze, etc..
	b	(reverse)
		X = strumento (instrument)
		Y = arnese (tool)
		Z = strumenti manuali (manual tools)
Effect:		strumenti NEAR_SYNONYM arnesi
		arnesi NEAR_SYNONYM strumenti

2.1.3 XPOS_NEAR_SYNONYMY

Cross-part of speech relations can be formulated using frames that explicitly compensate for the syntactic differences between the word pairs. The following tests express a synonymy relation between nouns and verbs in general:

Test 4

Comment: Synonymy of nouns and verbs denoting events and states

Score *Test sentence*

yes a If there is a case of a/an X then something/someone/it Ys

yes b If something/someone/it Ys then there is a case of a/an X

Conditions:

- X is a noun in the singular
- Y is a verb in the third person singular form
- there are no specifying PPs that apply to the X-phrase or the Y-phrase
- preferably there is a morphological link between the noun and the verb

Example: a If there is a case of a movement then something moves

b If something moves then there is a case of a movement

Effect: movement N XPOS_NEAR_SYNONYM move V
move V XPOS_NEAR_SYNONYM movement N

Test 4 - Dutch

Score *Test sentence*

yes a Als er sprake is van een X dan Yt er iets/iemand/het

yes b als iets/iemand/het Yt dan is er sprake van een X

Conditions: same

Example: a Als er sprake is van een beweging dan beweegt er iets

b als iets beweegt dan is er sprake van een beweging

X = beweging (movement)

Y = beweegt (move)

Effect: beweging N XPOS_NEAR_SYNONYM bewegen V
bewegen V XPOS_NEAR_SYNONYM beweging N

Test 4 - Spanish

Score *Test sentence*

yes a Si se da el caso de (un) X entonces es que algo/alguien (se) Y

yes b Si algo/alguien (se) Y entonces es que se da el caso de (un) X

Conditions:	same
Example:	a Si se da el caso de un movimiento entonces es que algo se mueve
	b Si algo se mueve entonces se da el caso de un movimiento
	X = movimiento (movement)
	Y = mover (move)
Effect:	movimiento N XPOS_NEAR_SYNONYM mover V
	mover V XPOS_NEAR_SYNONYM movimiento N

Test 4 -Italian

Score	Test Sentence
yes	a se c'è un X allora qualcosa/ qualcuno Y
yes	b se qualcosa/ qualcuno Y allora c'è un X
Conditions:	same
Example:	a se c'è una partenza allora qualcuno parte
	b se qualcuno parte allora c'è una partenza
	X = partenza (departure)
	Y = partire (to leave)
Effect:	partenza N XPOS_NEAR_SYNONYM partire V
	partire V XPOS_NEAR_SYNONYM partenza N

The distinction between hyponymy and synonymy is not always clear-cut. Sometimes concepts can be very close showing only a very limited specialization. In the case of relations across part-of-speech we can at least formulate the extra conditions that a strong morphological link between the two words is preferred, which is the case of *beweging* (movement) and *bewegen* (move).

Whereas the previous test works both for non-dynamic states and dynamic events, the next test only applies to dynamic events:

Test 5

<i>Comment:</i>	Synonymy between event denoting nouns and verbs
<i>Score</i>	<i>Test sentence</i>
yes	a if an X takes place then something/somebody/it Ys
yes	b if something/somebody/it Ys then a/an X takes place
<i>Conditions:</i>	- X is a noun in the singular
	- Y is a verb in the third person singular form
	- there are no specifying PPs that apply to the X-phrase or the Y-phrase
	- preferably there is a morphological link between the noun and the verb
<i>Example:</i>	X = movement
	Y = move
<i>Effect:</i>	movement N XPOS_NEAR_SYNONYM move V
	move V XPOS_NEAR_SYNONYM movement N

Test 5 - Dutch

Score	Test sentence
yes	a als er een X plaatsvindt dan Yt er iets/iemand/het
yes	b als iets/iemand/het Yt dan vindt er een X plaats
Conditions:	same
Example:	a als er een beweging plaatsvindt dan beweegt er iets/iemand

b als iets/iemand beweegt dan vindt er een beweging plaats
 X = beweging (movement)
 Y = beweegt (move)
 Effect: beweging N XPOS_NEAR_SYNONYM bewegen V
 bewegen V XPOS_NEAR_SYNONYM beweging N

Test 5 -Spanish

Not used.

Test 5-Italian

Not used.

Test 6

Comment: Synonymy between state-denoting nouns and verbs

Score *Test sentence*

yes a If there is a state of X then something/someone/it Ys

yes b If something/something/it Ys then there is a state of a/an X applies

Conditions:

- X is a noun in the singular
- Y is a verb in the third person singular form
- there are no specifying PPs that apply to the X-phrase or the Y-phrase
- preferably there is a morphological link between the noun and the verb

Example:

yes a If there is a state of sleep then something/someone/it sleeps

yes b If something/something/it sleeps then there is a state of a/an sleep applies

Effect: sleep N XPOS_NEAR_SYNONYM sleep V
 sleepV XPOS_NEAR_SYNONYM sleep N

Test 6 -Dutch

Score		Test sentence
yes	a	Als er sprake is van een toestand van X dan Yt iets/iemand/het
yes	b	als iets/iemand/het Yt dan is er sprake van een toestand van X
Conditions:		same
Example 1:	a	Als er sprake is van een toestand van slaap dan slaapt er iets
	b	als iets slaapt dan is er sprake van een toestand van slaap
		X = slaap (sleep)
		Y = slapen (sleeps)
Effect:		slaap N XPOS_NEAR_SYNONYM slapen V
		slapen V XPOS_NEAR_SYNONYM slaap N
Example 2:	a	Als er sprake is van een toestand van gelijkenis dan lijkt er iets
	b	als iets lijkt dan is er sprake van een toestand van gelijkenis
		X = gelijkenis (equality, similarity)
		Y = lijkt (equals, is similar)
Effect:		gelijkenis N XPOS_NEAR_SYNONYM lijken V
		lijken V XPOS_NEAR_SYNONYM gelijkenis N

Test 6 -Spanish

Score		Test sentence
yes	a	Si existe un estado de X entonces es que algo/alguien Y
yes	b	Si algo/alguien Y entonces es que existe un estado de X
Conditions:		same
Example:	a	Si existe un estado de sueño entonces es que alguien duerme
	b	Si alguien duerme entonces es que existe un estado de sueño
		X = sueño (sleep)
		Y = dormir (sleep)
Effect:		sueño N XPOS_NEAR_SYNONYM dormir V
		dormir V XPOS_NEAR_SYNONYM sueño N

Test 6-Italian

Score		Test sentence
yes		a se c'è uno stato di X allora qlco/qlcu Y
yes		b se qlco/qlcu Y allora si ha uno stato di X
Conditions:		same but there is a morphological link between noun and verb.
Examples:	a	se c'è uno stato di veglia allora qlco/qlcu veglia
	b	se qlco/qlcu veglia allora c'è uno stato di veglia
		X= veglia (watch)
		Y= vegliare (to stay up)
Effect:		veglia N XPOS_NEAR_SYNONYM vegliare V
		vegliare V XPOS_NEAR_SYNONYM veglia N

The next test is similar to the previous one but applies to adjectives/adverbs and nouns that denote non-dynamic states. The test is only different in so far that adjectives/adverbs need a copula to occur in the same sentence as above:

Test 7

Comment: Synonymy between state-denoting nouns and adjectives/adverbs

<i>Score</i>		<i>Test sentence</i>
yes	a	If there is a state of X then something/someone/it is Y
yes	b	If something/someone/it is Y then there is a state of a/an X applies
<i>Conditions:</i>		- X is a noun in the singular - Y is an adjective - there are no specifying PPs that apply to the X-phrase or the Y-phrase - preferably there is a morphological link between the noun and the adjective
<i>Example:</i>	a	If there is a state of carelessness then something/someone/it is careless
	b	If something/someone/it is careless then there is a state of a/an carelessness applies
<i>Effect:</i>		carelessness N XPOS_NEAR_SYNONYM careless A careless A XPOS_NEAR_SYNONYM carelessness N

Test 7 -Dutch

<i>Score</i>		<i>Test sentence</i>
yes	a	Als er sprake is van een toestand van X dan is iets/iemand/het Y
yes	b	als iets/iemand/het Y is dan is er sprake van een toestand van X
<i>Conditions:</i>		same
<i>Example 1:</i>	a	Als er sprake is van een toestand van onvoorzichtigheid dan is iets onvoorzichtig
	b	als iets onvoorzichtig is dan is er sprake van een toestand van onvoorzichtigheid X = onvoorzichtigheid(carelessness) Y = onvoorzichtig (careless)
<i>Effect:</i>		onvoorzichtigheid N XPOS_NEAR_SYNONYM onvoorzichtig A onvoorzichtig A XPOS_NEAR_SYNONYM onvoorzichtigheid N
<i>Example 2:</i>	a	Als er sprake is van een toestand van duisternis dan is het donker
	b	als het donker is dan is er sprake van een toestand van duisternis X = duisternis (darkness) Y = donker (dark)
<i>Effect:</i>		duisternis N XPOS_NEAR_SYNONYM donker A donker A XPOS_NEAR_SYNONYM duisternis N
<i>Example 3:</i>	a	Als er sprake is van een toestand van angst dan is iets/iemand bang
	b	als iets/iemand bang is dan is er sprake van een toestand van angst X = angst (fear) Y = bang (afraid)
<i>Effect:</i>		angst N XPOS_NEAR_SYNONYM bang A bang AXPOS_NEAR_SYNONYM angst N

Here we see in the first example that the noun "onvoorzichtigheid" (carelessness) has a straightforward morphological relation with "onvoorzichtig" (careless). However, the other two examples show that this is not a necessary condition.

Test 7 - Spanish

<i>Score</i>		<i>Test sentence</i>
yes	a	Si existe un estado de X entonces es que algo/alguien es/está Y
yes	b	Si algo/alguien es/está Y entonces es que existe un estado de X

Conditions:	- same
Example:	X = oscuridad (darkness) Y = oscuro (dark)
Effect:	oscuridad N XPOS_NEAR_SYNONYM oscuro A oscuro A XPOS_NEAR_SYNONYM oscuridad N

Test 7 - Italian

Score		Test sentence
yes	a	se c'è uno stato di X allora qlco/qlcu é Y
yes	b	se qlco/qlcu é Y allora si verifica uno stato di Y
Conditions:		same
Example:	a	se c'è uno stato di attività qlco/qlcu é attivo
	b	se qlco/qlcu é attivo c'è uno stato di attività
		X = attività (activity) Y = attivo (active)
Effect:		attività N XPOS_NEAR_SYNONYM attivo A attivo A XPOS_NEAR_SYNONYM attività N

2.2 Hyperonymy and Hyponymy

These are inverse relations, which roughly correspond to the notion of class-inclusion: if Y is a kind of X, then X is hyperonym of Y and Y is an hyponym of X. Both relations are as well asymmetric and transitive. A hyponymy relation implies that the hyperonym (the more general class) may substitute the hyponym (the more specific subtype) in a referential context but not the other way around. A referential context is a context where only the denotational range (the set of discourse entities) is considered (grammatical, register, pragmatic and other non-semantic properties of the considered words or context are neglected). Given these constraints there must be a full inclusion of the set of entities denoted by the hyponym in the set of entities denoted by the hyperonym. An extra constraint can be that there must be multiple co-hyponyms to result in a genuine hyponymy relation. This means that the denotation of the hyponym is never equal to the denotation of the hyperonym, i.e. it must be a proper subset. The same substitution principle as discussed above for synonymy can thus be applied to hyponymy relations but it only holds in one direction. However, to more clearly elicit the difference in specificity the tests have been extended with general specifying phrases. In addition to the formal substitution-sentences we can state that:

If a pair of words W1 and W2 fits (where the degree *yes---unclear---no* should be positive) the test frame then there should be at least one other word W3 which fits this frame in relation to W2 so that W1 and W3 are hyponyms of W2. The presence of co-hyponyms is a necessity to establish a genuine hyponymy relation.

In the next test three different paraphrases are used which elicit the implicational relation between the hyponym and the hyperonym.

2.2.1 HAS_HYPONYM and HAS_HYPERONYM

Test 8

<i>Comment:</i>		Hyponymy-relation between nouns
<i>Score</i>		<i>Test sentence</i>
yes	a	A/an X is a/an Y with certain properties) It is a X and therefore also a Y If it is a X then it must be a Y
no	b	the converse of any of the (a) sentences.
<i>Conditions:</i>		- both X and Y are singular nouns or plural nouns
<i>Example:</i>	a	Aa car is a vehicle with certain properties
	b	?A vehicle is a car with certain properties
	a	It is a car and therefore also a vehicle
	b	?It is a vehicle and therefore also a car
	a	If it is a car then it must be a vehicle
	b	?If it is a vehicle then it must be a car
<i>Effect:</i>		car N HAS_HYPERONYM vehicle N vehicle N HAS_HYPONYM car N

Test 8 - Dutch

<i>Score</i>		<i>Test sentence</i>
yes	a	een X is een Y met bepaalde eigenschappen het is een X en dus is het ook een Y als het een X is dan is het een Y
no	b	the converse of any of the a) sentences.
<i>Conditions:</i>		same
<i>Example:</i>	a	een auto is een voertuig met bepaalde eigenschappen
	b	?een voertuig is een auto met bepaalde eigenschappen
	a	het is een auto en dus is het ook een voertuig
	b	?het is een voertuig en dus is het ook een auto
	a	als het een auto is dan is het een voertuig
	b	?als het een voertuig is dan is het een auto
		X = auto (car)
		Y = voertuig (vehicle)
<i>Effect:</i>	auto N	HAS_HYPERONYM voertuig N
	voertuig N	HAS_HYPONYM auto N

Test 8 -Spanish

Score		Test sentence
yes	a	Un X es un Y con ciertas propiedades Es un X, y por consiguiente es un Y Si es un X necesariamente es un Y
no	b	the converse of any of the a) sentences.
Conditions:		- same
Example:	a	Un coche es un vehículo con ciertas propiedades
	b	?Un vehículo es un coche con ciertas propiedades
	a	Es un coche, y por consiguiente es un vehículo
	b	*Es un vehículo, y por consiguiente es un coche
	a	Si es un coche necesariamente es un vehículo
	b	*Si es un vehículo necesariamente es un coche
		X = coche (car) Y = vehículo (vehicle)
Effect:		coche N HAS_HYPERONYM vehículo N vehículo N HAS_HYPONYM coche N

Test 8 - Italian

Score		Test sentence
yes	a	Un X é un Y con certe proprietá É un X e perció é anche un Y Se é un X allora deve essere un Y
no	b	the converse of any of the (a) sentences
Conditions:		same
Examples:	a	uno strudel é un dolce con certe proprietá é uno strudel perció é un dolce se é uno strudel deve essere un dolce
	b	* un dolce é uno strudel con certe proprietá * é un dolce perció é uno strudel * se é un dolce deve essere uno strudel
		X = strudel Y = dolce (cake)
Effect:		strudel N HAS_HYPERONYM dolce N dolce N HAS_HYPONYM strudel N

Without the specifying phrase, this test can also be used for synonymy. The next test indicates a more specific type of hyponymy between kinds, species, races and brands:

Test 9

Comment: Hyponymy-relation between nouns of species and classes, which is reflected by the explicit hyponymy nouns such as sort/kind, type, race, species.

Score *Test sentence*
 yes a A/an X is a kind/type/race/species/brand of Y(s)

no b the converse of the (a) sentence.

Conditions:
 - X is a singular noun
 - Y is a singular or plural noun

Example: a A car is a kind of vehicle
 b ?A vehicle is a kind of car

Effect: car N HAS_HYPERONYM vehicle N
 vehicle N HAS_HYPONYM car N

This test cannot be used for synonymy.

Test 9 -Dutch

Comment: Two variations of the (a) sentence may apply

Score *Test sentence*
 yes a een X is een soort/type/ras/geslacht/merk Y(s)
 (a/an X is a kind/type/race/species/brand of Y(s))
 een auto is een Ytype/Ysoort/Yras/Ygeslacht/Ymerk
 (a/an X is a/an Y-kind/Y-type/Y-sort/Y-race/Y-species/Y-brand)

no b the converse of any of the (a) sentences.

Conditions:
 - same

Example: a een auto is een soort voertuig (a car is a kind of vehicle)
 b ?een voertuig is een soort auto (a vehicle is a kind of car)
 a een auto is een voertuigtype (a car is a vehicle-kind)
 b ?een voertuig is een autotype/merk (a vehicle is a car-kind)

X = auto (car)
 Y = voertuig (vehicle)
Effect: auto N HAS_HYPERONYM voertuig N
 voertuig N HAS_HYPONYM auto N

Test 9 -Spanish

Score *Test sentence*
 yes a Un X es un/a tipo/clase/especie de Ys
 no b the converse

Conditions:
 - same

Example: X = coche (car)
 Y = vehículo (vehicle)

Effect: coche N HAS_HYPERONYM vehículo N
 vehículo N HAS_HYPONYM coche N

Test 9 -Italian

Score		Test sentence
yes	a	Un X é un tipo, sorta, genere, razza, specie di Y
no	b	the converse
Conditions:		same
Examples:	a	Un bassotto é una specie di cane
	b	*Un cane é una specie di bassotto
		X = bassotto (dachshund)
		Y = cane (dog)
Effect:		bassotto N HAS_HYPERONYM cane N
		cane N HAS_HYPONYM bassotto N

2.2.2 HAS_XPOS_HYPERONYM and HAS_XPOS_HYPONYM

As in the case of near_synonymy, hyperonymy-hyponymy can also be established between words belonging to a different part of speech. This relation also come in inverse pairs. In section 2.1 we have seen some test sentences for synonymy-relations across parts-of-speech. In principle these tests can also be used as a basis for hyponymy-tests with some additions to elicit the difference in specificity:

Test 10

<i>Comment:</i>		Hyponymy of nouns and verbs denoting events and states
<i>Score</i>		<i>Test sentence</i>
yes/no	a	If there is a case of a/an X then something/someone/it Ys + NP, PP (in a certain way)
no/yes	b	If something/someone/it Ys then there is a case of a/an certain X
<i>Conditions:</i>		- X is a noun in the singular
		- Y is a verb in the third person singular form
		- there should be at least one specifying NP or PP that makes the Y-phrase equivalent to the X-phrase or the other way around.
		- preferably there is no morphological link between the noun and the verb
<i>Effect- 1</i>		X-N HAS_XPOS_HYPERONYM Y-V
		Y-V HAS_XPOS_HYPONYM X- N
<i>Effect- 2</i>		X-V HAS_XPOS_HYPERONYM Y-N
		Y-N HAS_XPOS_HYPONYM X-V

The reverse of the score leads to a reversion of the hyponymy as well: noun-to-hyperonym-verb or verb-to-hyperonym-noun. As long as one direction has a clear positive score and the other direction has a clear negative score we are dealing with a hyponymy relation. This will be more clearly noticed in the following language-specific examples:

Test 10 - Dutch

Score		Test sentence	
yes/no	a	als er sprake is van een X dan Yt (er) iets/iemand/het + NP, PP (op een bepaalde wijze)	
no/yes	b	als iets/iemand/het Yt dan is er sprake van een bepaalde X	
Conditions:		- same	
Example 1:	a	als er sprake is van een moord dan doodt iets/iemand/het illigaal	
	b	?als iets/iemand/het doodt dan is er sprake van een bepaalde moord	
		X = moord (murder)	
		Y = doodt(kills)	
Effect:		moord N HAS_XPOS_HYPERONYM doden V	
		doden V HAS_XPOS_HYPONYM moord N	
Example 2:	a	?als er sprake is van een sport dan voetbalt er iets op een bepaalde wijze	
	b	als iets/iemand/het voetbalt dan is er sprake van een bepaalde sport	
		X = sport (sport)	
		Y = voetbalt (plays football)	
Effect:		sport N HAS_XPOS_HYPONYM voetballen V	
		voetballen V HAS_XPOS_HYPERONYM sport N	

Test 10 - Spanish

Score		Test sentence	
yes/no	a	Si se da el caso de un X entonces es que alguien/algo Y	
no/yes	b	Si algo/alguien Y entonces se da el/un caso de (un cierto) X	
Conditions:		- same	
Example 1:	a	Si se da el caso de una caza entonces es que alguien mata	
	b	?Si alguien mata entonces se da un caso de caza	
		X = caza (hunt)	
		Y = matar (to kill)	
Effect:		caza N HAS_XPOS_HYPERONYM matar V	
		matar V HAS_XPOS_HYPONYM caza N	
Example 2:	a	?Si se da el caso de un deporte entonces es que alguien pesca	
	b	Si alguien pesca entonces se da el caso de un cierto deporte	
		X = deporte (sport)	
		Y = pescar (to fish)	
Effect:		deporte N HAS_XPOS_HYPONYM pescar V	
		pescar V HAS_XPOS_HYPERONYM deporte N	

Test 10- Italian

Score		Test sentence
yes/no	a	se si dá il caso di un X questo comporta che qlco/qlcu Ys + NP, PP (in un certo modo)
no/yes	b	se qlco/qlcu Ys comporta / si dá il caso di un X di qualche tipo
Conditions:		same
Example 1:	a	se si dá il caso di una caccia questo comporta che qualcuno uccida
	b	*se qualcuno uccide si dá il caso di una caccia
		X= caccia (hunt)
		Y= uccidere (to kill)
Effect:		caccia N HAS_XPOS_HYPERONYM uccidere V uccidere V HAS_XPOS_HYPONYM caccia N
Example 2:	a	*se si dá il caso che qualcuno faccia sport ciò comporta che qualcuno cacci
	b	se qualcuno caccia comporta un certo sport
		X = sport
		Y = cacciare (to hunt)
Effect:		sport N HAS_XPOS_HYPONYM cacciare V cacciare V HAS_XPOS_HYPERONYM sport N

The standard construction for hyponymy can be extended with a manner-specification to elicit troponymy relations between nouns and verbs:

Test 11

<i>Comment:</i>		Manner-hyponymy or troponymy between nouns and verbs
<i>Score</i>		<i>Test sentence</i>
yes/no	a	(a/an/the) X is a way/manner of Y
no/yes	b	Y is a certain X
<i>Conditions:</i>		- X is a noun in the singular - Y is a verb in the infinitive form - preferably there is no morphological link between the noun and the verb
<i>Effect 1:</i>		X- N HAS_XPOS_HYPERONYM Y- V Y- V HAS_XPOS_HYPONYM X- N
<i>Effect 2:</i>		X-V HAS_XPOS_HYPERONYM Y-N Y-N HAS_XPOS_HYPONYM X-V

Test 11 -Dutch

Score	Test sentence	
yes/no a	(een/het) X is een wijze/manier van Y	
no/yes b	Y is een bepaalde X	
Conditions:	- same	
Example 1: a	voetbal is een manier van sporten	
b	?sporten is een bepaald voetbal	
	X = voetbal (football)	
	Y = sporten (sport)	
Effect:	voetbal N HAS_XPOS_HYPERONYM	sporten V
	sporten V HAS_XPOS_HYPONYM	voetbal N
Example 2: a	?een sport is een manier van voetballen	
b	voetballen is een bepaalde sport	
	X = sport (sport)	
	Y = voetballen (play football)	
Effect:	sport N HAS_XPOS_HYPONYM	voetballen V
	voetballen V HAS_XPOS_HYPERONYM	sport N

Test 11 -Spanish

Score	Test sentence	
yes/no a	(el/un) X es una manera de Y	
no/yes b	Y es un cierto X	
Conditions:	- same	
Example 1: a	la caza es una manera de matar	
b	*matar es una cierta caza	
	X = caza (hunt)	
	Y = matar (to kill)	
Effect:	caza N HAS_XPOS_HYPERONYM	matar V
	matar V HAS_XPOS_HYPONYM	caza N
Example 2: a	*el deporte es una manera de cazar	
b	cazar es un cierto deporte	
	X = deporte (sport)	
	Y = matar (to kill)	
Effect:	deporte N HAS_XPOS_HYPONYM	cazar V
	cazar V HAS_XPOS_HYPERONYM	deporte N

Test 11 - Italian

Score		Test sentence
yes/no	a	X é un modo di Y
no/yes	b	Y é un certo X
Conditions:		same
Example 1:		a la corsa é un modo di muoversi b *muoversi é una certa corsa X= corsa (run, race) Y= muoversi (to move)
Effect:		corsa N HAS_XPOS_HYPERONYM muoversi V muoversi V HAS_XPOS_HYPONYM corsa N
Example 2:		a *lo sport é un modo di correre b correre é un certo sport X = sport Y = correre (to run)
Effect:		sport N HAS_XPOS_HYPONYM correre V correre V HAS_XPOS_HYPERONYM sport N

Whereas the previous tests works both for non-dynamic states and dynamic events, the next test only applies to dynamic events:

Test 12

<i>Comment:</i>		Hyponymy between event denoting nouns and verbs
<i>Score</i>		<i>Test sentence</i>
yes/no	a	if (a/an/the) X takes place then something/somebody/it Ys + NP, PP (in a certain way)
no/yes	b	if something/somebody/it Ys then (a/an/the) certain X takes place
<i>Conditions:</i>		- X is a noun in the singular. - Y is a verb in the third person singular form. - there should be at least one specifying NP or PP that makes the Y-phrase equivalent to the X-phrase or the other way around. - preferably there is no morphological link between the noun and the verb
<i>Effect:</i>		X- N HAS_XPOS_HYPONYM Y- V Y-V HAS_XPOS_HYPERONYM X-N

Test 12 -Dutch

Score		Test sentence	
yes/no	a	als er (een/het) X plaatsvindt dan Yt (er) iets/iemand/het + NP, PP (op een bepaalde wijze)	
no/yes	b	als iets/iemand/het Yt dan vindt er (een/het) bepaald(e) X plaats	
Conditions:		- same	
Example:	a	als er gemompel plaatsvindt dan spreekt er iets/iemand op een bepaalde manier	
	b	?als iets/iemand spreekt dan vindt er bepaald gemompel plaats X = gemompel (murmuring) Y = spreekt (speaks)	
Effect:		gemompel N HAS_XPOS_HYPONYM spreken V	
		spreekenV HAS_XPOS_HYPERONYM gemompel N	

Test 12 -Spanish

Score		Test sentence	
yes/no	a	Si tiene lugar (un/el) X, entonces (algo/alguien) Y + NP, PP (en cierta forma)	
no/yes	b	Si (algo/alguien) Y entonces tiene lugar (un/el) X.	
Conditions:		- X is a noun in the singular. - Y is a verb in the third person singular form. - there should be at least one specifying NP or PP that makes the Y-phrase equivalent to the X-phrase or the other way around. - preferably there is no morphological link between the noun and the verb	
Example	a	Si tiene lugar un banquete, entonces alguien come para festejar algo.	
	b	Si alguien come, entonces tiene lugar un banquete X = banquete (feast) Y = comer (to eat)	
Effect:		banquete N HAS_XPOS_HYPONYM comer V	
		comer V HAS_XPOS_HYPERONYM banquete N	

Test 12 - Italian

Score		Test sentence
yes/no	a	se c'è un/una X allora qlco/qlcu Y + NP, PP in un certo modo
no/yes	b	se qlco/qlcu Y allora c'è un certo X
Conditions:		- X is a noun in the singular form - Y is a verb in the third person singular form - there should be at least one specifying NP or PP that makes the Y-phrase equivalent to the X-phrase or the other way around. - preferably there is no morphological link between the noun and the verb
Example :	a	se c'è un mormorio allora qualcuno parla a bassa voce
	b	*se qualcuno parla allora c'è un certo mormorio
	a	?se c'è un mormorio allora qualcuno parla
	b	se qualcuno parla a bassa voce c'è un certo mormorio
		X= mormorio (whispering) Y= parlare (to speak)
Effect:		mormorio N HAS_XPOS_HYPERONYM parlare V parlare V HAS_XPOS_HYPONYM mormorio N

The next test only applies to non-dynamic states expressed by nouns and verbs:

Test 13

<i>Comment:</i>		Hyponymy between state-denoting nouns and verbs
<i>Score</i>		<i>Test sentence</i>
yes/no	a	if there is a state of X then something/someone Ys + NP, PP (in a certain way)
no/yes	b	if someone/something/it Ys then a state of a/an certain X applies
<i>Conditions:</i>		- X is a noun in the singular - Y is a verb in the third person singular form - there should be at least one specifying NP or PP that makes the Y-phrase equivalent to the X-phrase or the other way around. - preferably there is no morphological link between the noun and the verb
<i>Effect:</i>		X-N HAS_XPOS_HYPERONYM Y-V Y-V HAS_XPOS_HYPONYM X-N

Test 13- Dutch

Score		Test sentence	
yes/no	a	als er sprake is van een toestand van X dan Yt (er) iets/iemand/het + NP, PP (op een bepaalde wijze)	
no/yes	b	als iets/iemand/het Yt dan is er sprake van een toestand van bepaalde X	
Conditions:		- same	
Example:	a	als er sprake is van een toestand van clubliefde dan houdt iemand van een club	
	b	?als iemand van iets houdt dan is er sprake van een toestand van bepaalde clubliefde X = clubliefde (love N for a club) Y = houden van (to love)	
Effect:		clubliefde N XPOS_HYPERONYM houden van V houden van V XPOS_HYPONYM clubliefde N	

Test 13 - Spanish

Score		Test sentence	
yes/no	a	Si se da un estado de X, entonces es que algo/alguien Y+ NP, PP (de un cierto modo)	
no/yes	b	Si alguien/algo Y entonces se da un estado de un cierto X	
Conditions:		same	
Example 1:	a	Si se da un estado de misantropma, entonces es que alguien odia a los hombres en general.	
	b	*Si alguien odia, entonces se da un estado de una cierta misantropma X= misantropma (misanthropy) Y= odiar (to hate)	
Effect:		misanthropia N HAS_XPOS_HYPERONYM odiar V odiar V HAS_XPOS_HYPONYM misantropia N	

Test 13 - Italian

Score		Test sentence	
yes/no	a	se c'e' uno stato di X allora qlco/qlcu Y + NP, PP in un certo modo	
no/yes	b	se qlco/qlcu Y allora c'e un qualche stato di X	
Conditions:		same	
Example 1:	a	se c'e' uno stato di misantropia allora qlco/qlcu odia gli uomini in generale	
	b	*se qlco/qlcu odia c'e' uno stato di misantropia X= misantropia (misanthropy) Y= odiare (to hate)	
Effect:		misanthropia N HAS_XPOS_HYPERONYM odiare V odiare V HAS_XPOS_HYPONYM misantropia N	

The next test elicits hyponymy between adjectives/adverbs and nouns that denote non-dynamic states:

Test 14

<i>Comment:</i>		Hyponymy between state-denoting nouns and adjectives
<i>Score</i>		<i>Test sentence</i>
yes/no	a	if there is a state of X then something/someone/it in particular is Y in a certain way
no/yes	b	if something/someone/it is Y then a state of a/an X applies
<i>Conditions:</i>		- X is a noun in the singular - Y is an adjective - there is at least one specifying adverb, NP or PP that applies to the X-phrase or the Y-phrase - preferably there is a no morphological link between the noun and the adjective
<i>Effect:</i>		X-N HAS_XPOS_HYPERONYM Y-A Y-A HAS_XPOS_HYPONYM X-N

Test 14- Dutch

<i>Score</i>		<i>Test sentence</i>
yes/no	a	Als er sprake is van een toestand van X dan is iets/iemand/het (specifiek) ook Y
no/yes	b	als iets/iemand/het Y is dan is er sprake van een bepaalde toestand van X
<i>Conditions:</i>		- same
<i>Example:</i>	a	Als er sprake is van een toestand van verduistering dan is iets specifiek (hemellichamen = planets, stars, moons) ook donker
	b	?als iets donker is dan is er sprake van een bepaalde toestand van verduistering X = verduistering (eclipse) Y = donker (dark)
<i>Effect:</i>		verduistering N HAS_XPOS_HYPERONYM donker A donker A HAS_XPOS_HYPONYM verduistering N

Test 14 - Spanish

<i>Score</i>		<i>Test sentence</i>
yes/no	a	Si se da un estado de X entonces algo/alguien en particular es/está Y de algún modo
no/yes	b	Si alguien/algo es/está Y entonces se da el caso de que se da un estado de X
<i>Conditions:</i>		- same
<i>Example:</i>		X = eclipse (eclipse) Y = oscuro (dark)
<i>Effect:</i>		eclipse N HAS_XPOS_HYPERONYM oscuro A oscuro A HAS_XPOS_HYPONYM eclipse N

Test 14 - Italian

Score		Test sentence
yes/no	a	se c'è uno stato di X allora qlco/qlcu é Y in un certo modo
no/yes	b	se qlco/qlcu é Y allora c'e un qualche stato di X
Conditions:		same
Example 1:	a	se c'è un'eclisse il cielo é scuro in un certo modo
	b	*se il cielo é scuro c'è un'eclisse
		X= eclisse (eclipse)
		Y= scuro (dark)
Effect:		eclisse N HAS_XPOS_HYPERONYM scuro A
		scuro A HAS_XPOS_HYPONYM eclisse N

2.3 Antonymy

Intuitively, Antonymy is the relation which stands between lexical opposites, such as “to ascend” and “to descend”, “good” and “bad” or “justice” and “injustice”. It is clear that antonymy positively is a symmetric relation, but little more seems possible to be said neither about its formal status nor about how to define it, since it seems to encode a large range of phenomena of opposition. Miller and Fellbaum (1992), for instance, point out that *the antonym of a word X is sometimes not-X, but not always..* For example, “rich” and “poor” are antonyms, but to say that someone is not rich does not mean that they must be poor”.

It is also unclear whether Antonymy stands between either word forms or word meanings. For instance, “appearance” and “arrival” are, in the appropriate senses, synonyms; but linguistic intuition says that the appropriate antonyms are different for each word (“disappearance” and “departure”). With respect to this, EWN will assume the solution adopted by Miller's WordNet, that is, antonymy is considered to be a relation between word forms, but not between word meanings -namely synsets. Therefore, in the example above, the antonymy relation will hold between “appearance” and “disappearance”, “arrival” and “departure” as word forms. In those cases that antonymy also holds for the other variants of the synset we use a separate relation NEAR_ANTONYM. Finally, we may find cases in which there is an opposition between synsets with different parts-of-speech. Just as with the synonymy and hyponymy relations we store these relations as XPOS_NEAR_ANTONYM relations. Since it is not yet clear whether we will be able to code many examples for these we have not specified explicit tests for these. However, such tests can be easily derived in the same way as above by compensating for the difference in part-of-speech.

Antonyms typically form contrasting categories within the same dimension. This means that an Antonym not only contrasts with another Antonym in one or more features (e.g. animate/inanimate) but that they have to share the same hyperonym: i.e. they have to be competitors within a reasonable denotational range. This latter criterion prevents us from contrasting irrelevant pairs such as “car” and “love”. An antonymy test therefore has to consist of two parts: one part expressing the contrast and one part expressing the shared dimension or hyperonym

Test 15*Comment:* Antonymy of nouns*Score* *Test sentence*

yes a X and Y are both a kind of Z but X is the opposite of Y

yes b the converse of (a)

Conditions:
- X and Y are singular or plural nouns
- Z is a hyperonym of both X and Z and within a reasonable, competitive denotational range.*Example:* a man and woman are both a kind of human being but man is the opposite of woman

b woman and man are both a kind of human being but woman is the opposite of man

Effect:
man- N ANTONYM woman-N
woman-N ANTONYM man-N**Test 15 -Dutch***Score* *Test sentence*

yes a X en Y zijn beide een soort/vorm van Z maar (een) X is het tegenovergestelde van (een) Y

(X and Y are both a kind of Z but X is the opposite of Y)

yes b the converse of (a)

Condition: the same*Example:* a X and Y zijn een soort levend wezen maar een dier is het tegenovergestelde van een plant.

(animal and plant are a kind of living being but an animal is the opposite of a plant)

b X and Y zijn een soort levend wezen maar een plant is het tegenovergestelde van een dier

(plant and animal are a kind of living being but a plant is the opposite an animal)

X = dier (animal)

Y = plant (plant)

Z = levend wezen (living being)

Effect: dier N ANTONYM plant N

plant N ANTONYM dier N

Test 15- Spanish*Score* *Test sentence*

yes a X y Y son tipos de Z pero X es lo contrario de Y

yes b the converse

Conditions: same*Example:* a Hombre y mujer son tipos de persona humana pero hombre es lo contrario de mujer*Effect:* hombre- N ANTONYM mujer-N

`mujer-N ANTONYM hombre-N

Test 15 - Italian

Score		Test sentence
yes	a	X e Y sono entrambi un tipo di Z, ma essere un X é l'opposto dell'essere un Y
yes	b	X e Y sono entrambi un tipo di Z, ma essere un Y é l'opposto dell'essere un X
Conditions:		same
Examples:	a	un uomo e una donna sono entrambi esseri umani, ma essere un uomo é l'opposto dell'essere una donna
	b	un uomo e una donna sono entrambi esseri umani, ma essere una donna é l'opposto dell'essere un uomo
		X = uomo (men)
		Y = donna (woman)
		Z = essere umano (human being)
Effect:		uomo N ANTONYM donna N
		donna N ANTONYM uomo N

2.4 Part-whole Relations

Most scholars working within both Lexical Semantics (e.g. Cruse, 1986) and Psycholinguistics (e.g. Winston et al. 1987) pose that the so-called Part-Whole relation is not merely a relation but a family of them, being the most salient: (i) that which stands between (the nouns standing for) a whole and their constituent parts (“part”, e.g. “hand”-“finger”); (ii) between a portion and the whole from which it has been detached (“portion”, e.g. “ingot”-“metal”); (iii) between a place and a wider place which includes it (“location”, e.g. “oasis”-“desert”); (iv) between a set and their members (e.g. “fleet” -“ship”); and (v) between a thing and the substance it is made of (“made-of”, e.g. “book”-“paper”).

Undoubtedly more kinds of part-whole relations can be stated, and probably some relations that we can retrieve from our sources will not easily suit those mentioned above. For these reasons it has been decided: (i) to limit part-whole relations to be encoded in EWN to those five types: part, portion, location, member, and made-of; and (ii) to let an underspecified relation to cover unclear cases.

Moreover, a general distinction can be made between unique and non-unique parts. Unique parts belong to one type of whole, e.g. *finger* which is only a part of *hand*, non-unique parts can belong to a diverse range of wholes, e.g. *window* which can be a part of a *building*, *vehicle*, *container*, etc.. Whether or not a part is unique follows from the fact that there are multiple disjunctive wholes to which it is linked.

Part-Whole relations clearly come in inverse pairs, namely holonymy and meronymy - if X is the holonym of Y, Y is the meronym of X. Consequently, we will define in EWN a general relation HAS_HOLONYM (and its inverse HAS_MERONYM) and five subtypes of them, namely HAS_HOLO_PART and HAS_MERO_PART, HAS_HOLO_PORTION and HAS_MERO_PORTION, HAS_HOLO_LOCATION and HAS_MERO_LOCATION,

HAS_HOLO_MEMBER and HAS_MERO_MEMBER, and HAS_HOLO_MADEOF and HAS_MERO_MADEOF.

As explained above, the automatically reversed relations will get the label **reversed**. In the examples below we will not express this because the tests do not make clear which direction of the relation was explicitly coded and which direction was the result of automatic reversal. However, when we extract this information from dictionaries there is often a clear direction in which a relation is stored. If a relation is extracted twice from opposite directions neither of the inverse relations will get the label **reversed**. Note that this can mean that a reversed relation may be first automatically generated and later on extracted explicitly, in which case the **reversed** label has to be removed again. Furthermore, the automatic reversal of a relation should be blocked when it is already present as an explicit relation.

2.4.1 HAS_MERONYM and HAS_HOLONYM

Test 16

<i>Comment:</i>	General meronymy for nouns	
<i>Score</i>	<i>Test sentence</i>	
yes	a	(a/an) X makes up a part of (a/an) Y (a/an)Y has (a/an) Xs
no	b	the converse of the a) relations
<i>Conditions:</i>	X and Y are concrete nouns and are interpreted generically	
<i>Effect:</i>	X	HAS_HOLONYM Y
	Y	HAS_MERONYM X

The examples given here can all be interpreted as more specific relations as well.

Test 16 - Dutch

<i>Score</i>	<i>Test sentence</i>	
yes	a	(een) X maakt deel uit van (een) Y (een) Y heeft (een) X
no	b	the converse of the a) relations
<i>Conditions:</i>	same	
<i>Example:</i>	a	een wiel maakt deel uit van een voertuig een voertuig heeft een wiel
	b	?een voertuig maakt deel uit van een wiel ?een wiel heeft een voertuig X = wiel (wheel) Y = voertuig (vehicle)
<i>Effect:</i>	wiel N	HAS_HOLONYM voertuig N
	voertuig N	HAS_MERONYM wiel N

Another general meronymy test for Dutch is based in the possessive-construction, although it can also be applied to other types of relations (e.g. kinship relations) and in some case the direction of the relation is not elicited:

Test 16.1 - Dutch

Comment: General meronymy for nouns using possessive constructions

Score *Test sentence*

yes a de/het X van (een) Y
(the X of (a/an) Y)
de/het Y z'n/haar X
(the Y his/her X)
de/het Ys X
(the Y's X)

yes b de/het Y van (een) X
yes de/het X z'n/haar Y
no de/het X's Y

Conditions: X and Y are concrete nouns

Example: a de leuning van een stoel

de stoel z'n leuning

de stoels leuning

b !de stoel van een leuning

!de leuning z'n stoel

?de leunings stoel

X = leuning (back)

Y = stoel (chair)

Effect: leuning N HAS_HOLONYM stoel N
stoel N HAS_MERONYM leuning N

Test 16 - Spanish

Comment: General meronymy for nouns

Score *Test sentence*

yes a un/el X es parte de un/el Y
un/el Y tiene (un/el) Xs

no b the converse of the a) relations

Conditions: same

Examples a un dedo es parte de una mano

b la mano tiene dedos

X = finger

Y = hand

Effect: dedo HAS_HOLONYM mano
mano HAS_MERONYM dedo

Test 16 -Italian

Score		Test sentence
yes	a	un X é parte di un Y un Y ha un X
no	b	the converse of the (a) relations
Conditions:		same
Examples:	a	il motore é parte della macchina
	b	la macchina ha un motore
		X= motore (engine) Y= macchina (car)
Effect:		motore N HAS_HOLONYM macchina N macchina N HAS_MERONYM motore N

2.4.2 HAS_HOLO_MEMBER and HAS_MERO_MEMBER

Test 17

<i>Comment:</i>		Member-group meronymy for nouns using a relational member-noun
<i>Score</i>		<i>Test sentence</i>
yes	a	(a/an) X is a member/element of (a/an/the) Y
no	b	the converse of a)
<i>Conditions:</i>		- X is a single object-denoting noun - Y is a multiform noun (either a group-noun, a collective-noun or as a lexicalized plural denoting multiple objects) - preferably humans, animals, plants or vehicles or closed sets such as the number system, or the alphabet.
<i>Effect:</i>		X HAS_HOLO_MEMBER Y Y HAS_MERO_MEMBER X

Test 17 - Dutch

Score		Test sentence
yes	a	(een) X is een lid/element van (een/de/het) Y
no	b	the converse of (a)
Conditions:		- same
Example:	a	een letter is een lid/element van het alfabet
	b	?het alfabet is een lid/element van de letter X = letter (character) Y = alfabet (alphabet)
Effect:		letter N HAS_HOLO_MEMBER alfabet N alfabet N HAS_MERO_MEMBER letter N
Example:	a	een auto is een lid/element van het verkeer
	b	?verkeer is een lid/element van een auto X = auto (car) Y = verkeer (traffic)
Effect:		auto N HAS_HOLO_MEMBER verkeer N verkeer N HAS_MERO_MEMBER auto N
Example:	a	een commissielid is een lid/element van een commissie
	b	?een commissie is een lid/element van een commissielid

X = commissielid (committee member)
 Y = commission (committee)

Effect: commissielid N HAS_HOLO_MEMBER commission N
 commission N HAS_MERO_MEMBER commissielid N

Test 17 - Spanish

Score Test sentence
 yes a (un) X es (miembro/un) elemento de (un) Y
 no b (the converse of a)
 Conditions: - same
 Example: X = cerdo (pig)
 Y = piara (herd of pigs)

Effect: cerdo HAS_HOLO_MEMBER piara
 piara HAS_MERO_MEMBER cerdo

Test 17 - Italian

Score Test sentence
 yes a un X é un membro / un elemento di Y
 no b the converse of the (a) relations
 Conditions: same
 Examples: a un/una deputato é un membro del parlamento
 X= deputato (deputy)
 Y= parlamento (parliament)

Effect: deputato N HAS_HOLO_MEMBER parlamento N
 parlamento N HAS_MERO_MEMBER deputato N

For Dutch, the converse test has to be differentiated for homogeneous groups (consisting of a single type of elements/members) and heterogeneous groups:

Test 17.1 - Dutch

Comment: Heterogeneous member-group meronymy for nouns using a relational group-noun

Score *Test sentence*
 yes a (een/het) X is een aantal/groep/verzameling/set/stel/paar/collectie/stapel/reeks/ rij van elementen/leden/delen waaronder Y
 ((a/an/the) X is a number/group/collection/set/set/pair/collection/stack/range/row of elements/members/parts among which Y)

no b the converse of (a)
Conditions:
 - X is a multiform noun (either a group-noun, a collective-noun or as a lexicalized plural denoting multiple objects) consisting of diverse elements/members.
 - Y is a single-object denoting noun
 - preferably humans, animals, plants or vehicles or closed sets such as the number system, or the alphabet.

Example: a het gereedschap is een verzameling van elementen waaronder een hamer.
 b ?een hamer is een verzameling van elementen waaronder gereedschap.
 X = gereedschap (tools)
 Y = hamer (hammer)

Effect: gereedschap N HAS_MERO_MEMBER hamer N
 hamer N HAS_HOLO_MEMBER gereedschap N

Test 17.2 Dutch

Comment: Homogeneous member-group meronymy for nouns using a relational group-noun

Score *Test sentence*

yes a (een/het) X is een aantal/groep/verzameling/set/
 stel/paar/collectie/stapel/reeks/rij van Ys

no b ((a/an/the) X is a
 number/group/collection/set/set/pair/collection/stack/range/row of Ys)
 the converse of (a)

Conditions: - X is a multiform noun (either a group-noun, a collective-noun or as a lexicalized plural denoting multiple objects) consisting of homogeneous elements/members.
 - Y is a single-object denoting noun in the plural form
 - preferably humans, animals, plants or vehicles or closed sets such as the number system, or the alphabet.

Example: a een file is een groep van autos.
 b ?een auto is een groep van files.
 X = file (tailback)
 Y = auto's (cars)

Effect: file N HAS_MERO_MEMBER auto N
 auto N HAS_HOLO_MEMBER file N non-factive

Example: a een commissie is een groep van commissieleden.
 b ?een commissielid is een groep van commissies.
 X = commission (committee)
 Y = commissieleden (committee members)

Effect: commission N HAS_MERO_MEMBER commissielid N
 commissielid N HAS_HOLO_MEMBER commission N

2.4.3 HAS_HOLO_PORTION and HAS_MERO_PORTION

Several studies have noticed that what can be called the *portion-of* relation differs in many aspects from other meronymic relations; mainly in that (i) the whole always pre-exist the portion; (ii) usually portions (as concepts) do not receive a differentiated lexical item to refer to them (for instance, there is no lexical item equivalent to “portion of cake”); (iii) boundaries of portions usually are not motivated; and (iv) portions usually lack a definite relation relative to their wholes. Although not frequent, sometimes portions are conventionalised enough in a particular language to give rise to a lexical unit. This lexical items will be linked to their wholes by means of a *has_holo-portion* link according to the following test:

Test 18

Comment: Portion meronymy for nouns using a relational amount-noun

Score *Test sentence*

yes a (a/an) X is an (amount/piece/portion) of Y

no b the converse of (a)
Conditions: X and Y are substance denoting nouns
Effect: Y HAS_MERO_PORTION X
 X HAS_HOLO_PORTION Y

Test 18 - Dutch

Score Test sentence
 yes a een X is een hoeveelheid/stuk/beetje Y
 no b the converse of (a)
Conditions: X and Y are substance denoting nouns
Example: a een slok is een hoeveelheid/stuk/beetje drank
 b ?een drank is een hoeveelheid/stuk/beetje slok
 X = slok (nip, sip)
 Y = drank (drink)
Effect: drank N HAS_MERO_PORTION slok N
 slok N HAS_HOLO_PORTION drank N

Test 18 - Spanish

Score Test sentence
 yes a un X es un trozo de (un) Y
 no b the converse of (a)
Conditions: X and Y are nouns
Example: a un lingote es un trozo de metal
 b *un metal es un trozo de un lingote
 X = lingote (ingot)
 Y = metal (metal)
Effect: metal N HAS_MERO_PORTION lingote N
 lingote N HAS_HOLO_PORTION metal N

Test 18 - Italian

Score		Test sentence
yes	a	un/una X é un pezzo, una porzione di Y
no	b	the converse of the (a) relations
Conditions:		X and Y are nouns.
Examples:	a	una fetta é una porzione di pane X= fetta (slice) Y= pane (bread)
Effect:		pane N HAS_MERO_PORTION fetta N fetta N HAS_HOLO_PORTION pane N

For Spanish also the following test may be required:

Test 18.1 - Spanish

Score		Test sentence
yes	a	(un/el) Y puede ser dividido en Xs (an Y may be divided/portioned in Xs)
no	b	the converse of (a)
Conditions:		X and Y are nouns
Example:	a	el metal puede ser dividido en lingotes
	b	*el lingote puede ser dividido en metales X = lingote (ingot) Y = metal (metal)
Effect:		metal N HAS_MERO_PORTION lingote N lingote N HAS_HOLO_PORTION metal N

2.4.4 HAS_HOLO_PART and HAS_MERO_PART

This relation stands for the typical component-whole relation, namely, something which is either topologically or temporally included in a larger entity and which as well bears some kind of autonomy, non-arbitrary boundaries, and a definite function with respect to the whole.

Test 19

<i>Comment:</i>		Complex meronymy for nouns
<i>Score</i>		Test sentence
yes	a	a/an X is a component of a/an Y
no	b	the converse of (a)
<i>Conditions:</i>		X and Y are concrete nouns denoting objects, there must be several Xs
<i>Effect:</i>		X HAS_HOLO_PART Y Y HAS_MERO_PART X

Test 19 - Dutch

Score		Test sentence
yes	a	een X is een onderdeel/component van een Y
no	b	the converse of (a)
Conditions:		same
Example:	a	een fietsbel is een onderdeel/component van een fiets
	b	?een fiets is een onderdeel/component van een fietsbel
		X = fietsbel (bicycle bell)
		Y = fiets (bicycle)
Effect:		fietsbel N HAS_HOLO_PART fiets N
		fiets N HAS_MERO_PART fietsbel N

Test 19 - Spanish

Score		Test sentence
yes	a	(un/el) X es un componente de (un/el) Y
no	b	the converse of (a)
Conditions:		same
Example:		X = motor (engine)
		Y = coche (car)
Effect:		motor HAS_HOLO_PART coche
		coche HAS_MERO_PART motor

Test 19 - Italian

Score		Test sentence
yes	a	un/il X é un componente di Y
no	b	the converse of the (a) relations
Conditions:		same
Examples:	a	la tastiera é un componente del computer
		X= tastiera (keyboard)
		Y= computer
Effect:		tastiera N HAS_HOLO_PART computer N
		computer N HAS_MERO_PART tastiera N

The condition claims that there must be multiple components (which can be of the same type) and that both the holonym and the meronym should be concrete objects. Complex holonyms can also contain substances but in that case the MADE_OF relation is used (see below). The converse test for a complex meronymy relations is the following:

Test 20

<i>Comment:</i>		Complex meronymy for nouns using a relation whole-noun
<i>Score</i>		<i>Test sentence</i>
yes	a	a/an Y is a whole/system/complex/network/arrangement/construction of parts/components among which a/an X
no	b	the converse of (a)
<i>Conditions:</i>		X and Y are concrete nouns denoting objects
<i>Effect:</i>		X HAS_HOLO_PART Y
		Y HAS_MERO_PART X

Test 20 - Dutch

Score		Test sentence
yes	a	een Y is een geheel/systeem/complex/netwerk/arrangement/constructie van delen/componenten/onderdelen waaronder een X
no	b	the converse of (a)
Conditions:		same
Example:	a	een fiets is een geheel/systeem/complex/netwerk/arrangement/constructie van onderdelen waaronder een ketting
	b	?een ketting is een geheel/systeem/complex/netwerk/arrangement/constructie van onderdelen waaronder een fiets X = ketting (chain) Y = fiets (bicycle)
Effect:		ketting N HAS_HOLO_PART fiets N fiets N HAS_MERO_PART ketting N

Test 20 - Spanish

Score		Test sentence
yes	a	el/un X es un todo/sistema/complejo/red/agrupación/constructo de partes/componentes, entre las cuales existe el/un Y
no	b	the converse of (a)
Conditions:		same
Example:		X = ordenador (computer) Y = teclado (keyboard)
Effect:		teclado HAS_HOLO_PART ordenador ordenador HAS_MERO_PART teclado

Test 20- Italian

Score		Test sentence
yes	a	un/il Y é una unità, sistema, complesso, rete.. di parti, componenti elementi fra cui un X
no	b	the converse of the (a) relations
Conditions:		same
Examples:	a	la bicicletta é una unità che ha come componente il manubrio Y= bicicletta (bicycle) X= manubrio (handlebars)
Effect:		manubrio N HAS_HOLO_PART bicicletta N bicicletta N HAS_MERO_PART manubrio N

2.4.5 HAS_HOLO_MADEOF and HAS_MERO_MADEOF

It seems that there are two basic ways of naming entities in the world, namely either as an individuated thing or as the stuff from which things are made of. This way, for instance a book can be alternatively named “a book” or “paper”. The relation which stands between things and the stuff which compose them is that which we term made-of. It is defined according to the suitability of the following test:

Test 21

<i>Comment:</i>		made-of meronymy for nouns		
<i>Score</i>		<i>Test sentence</i>		
yes	a	a/an X is made of Y		
no	b	the converse of (a)		
<i>Conditions:</i>		- X is a concrete object		
		- Y is a concrete substance		
<i>Effect:</i>		X	HAS_MERO_MADEOF	Y
		Y	HAS_HOLO_MADEOF	X

Test 21 - Dutch

<i>Score</i>		<i>Test sentence</i>		
yes	a	een X is gemaakt van Y		
no	b	the converse of (a)		
<i>Conditions:</i>		- same		
<i>Example:</i>	a	een tafel is gemaakt van hout		
	b	?een hout is gemaakt van tafel		
		X = tafel (table)		
		Y = hout (wood)		
<i>Effect:</i>		tafel N	HAS_MERO_MADEOF	hout N
		hout N	HAS_HOLO_MADEOF	tafel N

Test 21 - Spanish

<i>Score</i>		<i>Test sentence</i>		
yes	a	el/un X está hecho de Y		
no	b	the converse of (a)		
<i>Conditions:</i>		- same		
<i>Example:</i>		X = libro (book)		
		Y = papel (paper)		
<i>Effect:</i>		libro	HAS_MERO_MADEOF	papel
		papel	HAS_HOLO_MADEOF	libro

Test 21-Italian

Score		Test sentence
yes	a	un/il Y é fatto di X
no	b	the converse of the (a) relations
Conditions:		same
Examples:	a	il tavolo é fatto di legno Y= tavolo (table) X= legno (wood)
Effect:		tavolo HAS_MERO_MADEOF legno N legno N HAS_HOLO_MADEOF tavolo N

2.4.6 HAS_HOLO_LOCATION and HAS_MERO_LOCATION

Nouns of places are an important set within a lexical database. Since space, in a general sense, is by definition contiguous and sub-division in more inclusive pieces of space seems largely to be a matter of lexicalisation, it seems that nouns for places or locatum must stand in a relation of lexical-semantic inclusion to the nouns of the larger places which include them; a relation which is coherent to the topological 'real-world' relation which stands between the places named.

Test 22

<i>Comment:</i>		location meronymy for nouns
<i>Score</i>		<i>Test sentence</i>
yes	a	(a/an/the) X is a place located in (a/an/the) Y
no	b	the converse of (a)
<i>Conditions:</i>		- X is a concrete noun - Y is a concrete noun
<i>Effect:</i>		X HAS_HOLO_LOCATION Y Y HAS_MERO_LOCATION X

Test 22 - Dutch

Score		Test sentence
yes	a	(een/het) X is een plaats die zich bevindt in (een/het) Y
no	b	the converse of (a)
Conditions:		- same
Example:	a	het centrum is een plaats die zich bevindt in een stad
	b	?een stad is een plaats die zich bevindt in het centrum X = centrum (center) Y = stad (city)
Effect:		centrum N HAS_HOLO_LOCATION stad N stad N HAS_MERO_LOCATION centrum N

Test 22 - Spanish

Score		Test sentence
yes	a	un/el X es un/el lugar situado en un/el Y
no	b	the converse of (a)
Conditions:		- same
Example:		X = oasis (oasis) Y = desierto (desert)
Effect:		oasis HAS_HOLO_LOCATION desierto desierto HAS_MERO_LOCATION oasis

Test 22 - Italian

Score		Test sentence
yes	a	un/il X é un luogo situato in Y
no	b	il contrario di (a)
Conditions:		same
Examples:	a	una stazione é un luogo posto/situato in una città X= stazione (station) Y= città (town)
Effect:		stazione N HAS_HOLO_LOCATION città N città N HAS_MERO_LOCATION stazione N

2.5 “Role” and “Involved” relations and their subtypes.

From a cognitive point of view, function happens to be one of the major features which organise human knowledge. For instance, there seems to exist a tight conceptual relation between the action of reading and objects such as books, novels, newspapers, deliverables, and so. This relation of functionality is obviously reflected in the lexicon. Languages are rich in derivation procedures which straightforwardly generate nouns from verbs or the other way round -e.g. *run/runner*, *telephone/to telephone*. In such cases, there is a tight semantic relation between both lexical units that is potentially useful for any linguistic engineering task.

Such relations are mainly related to telicity but, since they also cover other aspects of semantic entailment, they will be referred to as - more generically - *involvement* relations. If the relation goes from nouns to verbs, it will be called *role*, being the inverse *involved*. The *role* relation is intended to encode data on a strict semantic relationship holding between a noun and a verb, with the verb referring to a (non-dynamic or dynamic) situation in which the entity indicated by the noun plays a role, semantically characterising the noun itself. Similarly, when the meaning of a verb is strongly characterised by a noun involved in its meaning, we encode such relation by means of the *involved* link. In many cases this relation is reflected by a morphological relation between noun and verb, as in *worker/work* and *hammer/to-hammer*. (See EWN deliverable 6, on the *Definition of links and subsets for verbs*, section 2.5, for details.)

The following roles have been distinguished:

ROLE_AGENT

ROLE_PATIENT
 ROLE_INSTRUMENT
 ROLE_LOCATION
 ROLE_DIRECTION
 ROLE_SOURCE_DIRECTION
 ROLE_TARGET_DIRECTION

As suggested above, for all undecidable cases which still have a strong role/involvement relation we keep a default ROLE & INVOLVED relation. The test for the general relation (which should apply to any of the above relations) is:

Test 23

Comment: Involvement/role relation in general
Score *Test sentence*
 yes (a/an) X is the one/that who/which is involved in Ying
Conditions: X is a noun
 Y is a verb in the infinitive form
Effect: X-N ROLE Y-V
 Y-V INVOLVED X-N reversed

Test 23 - Dutch

Score *Test sentence*
 yes (een) X is degene/dat die/wat bij het Y betrokken is
Conditions: same
Example: een hamer is dat wat bij het timmeren betrokken is
 X = hamer (hammer)
 Y = timmeren (hammer)
Effect: hamer N ROLE timmeren V
 timmeren V INVOLVED hamer N reversed

Test 23- Spanish

Score *Test sentence*
 yes (Un) X es (aquel/aquello) implicado en Y.
Conditions: X is a noun
 Y is a verb in the infinitive form
Effect: X-N ROLE Y-V
 Y-V INVOLVED X-N reversed

As the nearest spanish word for *involved* is not fully equivalent to this verb, an alternative - more complex - test can be considered:

Test 23bis -Spanish

<i>Comment:</i>	Involvement/role relation in general		
<i>Score</i>	<i>Test sentence</i>		
yes	a	un X es un X porque Y (an X is an X because it Ys)	
yes	b	un X es un X porque es una cosa que sirve para Y (an X is an X because it is a thing usually used to Y)	
yes	c	un X es un X porque puede ser Ydo/ha (sido) Ydo (an X is an X because it can be Yed/has (been) Yed)	
yes	d	un X es un X porque es un lugar que sirve para Y en el (an X is an X because it is a place which is usually used to Y in it)	
<i>Conditions:</i>	X is a noun Y is a verb at least one of the four sentences (a,b,c,d) must apply		
<i>Examples</i>	(1)	un corredor es un corredor porque corre (a runner is a runner because he runs)	
	(2)	un cuchillo es un cuchillo porque es una cosa que sirve para cortar (a knife is a knife because it is a thing usually used to cut)	
	(3)	un libro es un libro porque puede ser leído (a book is a book because it can be read)	
	(4)	un dormitorio es un dormitorio porque es un lugar que sirve para dormir en el (a bedroom is a bedroom because it is a place which is usually used to sleep in it)	
<i>Effects:</i>	corredor-N	ROLE	correr-V
	correr-V	INVOLVED	corredor-N reversed
	cuchillo-N	ROLE	cortarV
	cortar-V	INVOLVED	cuchilloN reversed
	libro-N	ROLE	leer-V
	leer-V	INVOLVED	libro-N reversed
	dormitorio-N	ROLE	dormir-V
	dormir-V	INVOLVED	dormitorio-N reversed

Test 23- Italian

<i>Score</i>	<i>Test sentence</i>		
yes	a	un/il X é ciò che/colui che é implicato in Y	
<i>Conditions:</i>	same		
<i>Example:</i>	a	il martello é usato per martellare X= martello (hammer) Y= martellare (to hammer)	
<i>Effect:</i>	martello N	ROLE	martellare V
	martellare V	INVOLVED	martello N

Specific subtypes of the role/involvement relation, namely involved-agent, involved-patient, involved-instrument, involved-location, involved-direction, involved-source-direction and involved-target-direction (and role reversals) will emerge from the specification of the test above in order to

include the notions of agent, patient, instrument, location, source-direction and target-direction. The corresponding tests are described in D006 -which is devoted to the verbal relations.

2.6 BE_IN_STATE & STATE_OF

This relation is needed encode links between nouns that refer to anything being in a particular state expressed by an adjective. These nouns often have an open denotation: i.e. they can refer to any entity to which the state applies, e.g. “the poor” refers to all entities which are in a “poor” state. Note that these nouns are not equivalent to the states, the entities which have the property “poor” are not states but normal first-order entities. The general test for this relation is:

Test 24

Comment:	being in a particular state		
Score	Test sentence		
yes a	a/an/the X is the one/that to whom/which the state Y applies		
Conditions:	X is a noun		
	Y is an adjective/adverb		
Example: a	the poor are the ones to whom the state poor applies		
	X = poor N (a poor person)		
	Y = poor A		
Effect:	poor N	BE_IN_STATE	poor A
	poor A	STATE_OF	poor N reversed

Test 24-Dutch

Comment:	being in a particular state		
Score	Test sentence		
yes a	(een) X is degene/datgene waarop de toestand Y betrekking heeft ((a/an) X is the one/that to whom/which the state Y applies)		
Conditions:	X is a noun		
	Y is an adjective/adverb		
Example: a	een arme is degene waarop de toestand arm betrekking heeft		
	X = arme (a poor person)		
	Y = arm (poor A)		
Effect:	arme N	BE_IN_STATE	arm A
	arm A	STATE_OF	arme N reversed

Test 24-Spanish

Comment:	being in a particular state		
Score	Test sentence		
yes	a	un/el X es aquel/aquello de quien/de lo cual es aplicable decir que se encuentra en estado 'Y'	
Conditions:	X is a noun Y is an adjective/adverb		
Example:	a	un pobre es aquel de quien es aplicable decir que se encuentra en estado 'pobre' X = pobre N (a poor person) Y = pobre A (poor A)	
Effect:	pobre N	BE_IN_STATE	pobre A
	pobre A	STATE_OF	pobre N reversed

Test 24-Italian

Comment:	being in a particular state		
Score	Test sentence		
yes	a	un/uno/una X e' un/uno/una a cui lo stato Y si applica	
Conditions:	X is a noun Y is an adjective / adverb		
Example:	a	un povero e' colui a cui si applica l'essere povero X= povero (poor person) Y= povero (poor A)	
Effect:	povero N	BE_IN_STATE	povero A
	povero A	STATE_OF	povero N

2.7 Instance / Class Relations

Hyperonymy-Hyponymy, defined above, is a relation which stands between classes of entities. But individual entities also belong to some class. Hence, relation `has_instance` (and its inverse `belongs_to_class`) have also been defined in the system:

Test 25

Comment:	Individuals belonging to a class		
Score	Test sentence		
yes	a	X is one of the Ys	
no	b	Y is one of the Xs	
Conditions:	X is a proper noun Y is a noun		
Example:	a	Manchester is one of the cities	
Effect:	Manchester	BELONGS_TO_CLASS	city
	city	HAS_INSTANCE	Manchester

Test 25 - Dutch

Score		Test sentence
yes	a	X is een van de Yen
no	b	Y is een van de Xen
Conditions:		X is a proper noun Y is a plural noun or substance noun
Example	a	Amsterdam is een van de steden
Effect:		Amsterdam BELONGS_TO_CLASS stad stad HAS_INSTANCE Amsterdam

Test 25- Spanish

Score		Test sentence
yes	a	X es uno/una de los/las Y
no	b	Y es uno/una de los/las X
Conditions:		same
Example:	a	Salamanca es una de las ciudades
Effect:		Salamanca BELONGS_TO_CLASS ciudad ciudad HAS_INSTANCE Salamanca

Test 25- Italian

Score		Test sentence
yes	a	X é una delle Y
no	b	Y é una delle X
Conditions:		same
Example:	a	Pisa é una delle città X= Pisa Y= città (town)
Effect:		Pisa Proper Noun BELONGS_TO_CLASS città N cittá N HAS_INSTANCE Pisa Proper Noun

2.8 Undefined Relations: “fuzzynyms”

Last, it has been decided to distinguish a rest relation to cover all the cases in which a word should be related to another one to be a part of the net but the relation has not been defined yet. Fuzzynymy will hold when all tests above fail but the test *X has some strong relation to Y* still works. A fuzzynum relation can hold between words with the same part-of-speech (FUZZYNYM) or across part-of-speech (XPOS_FUZZYNYM).

2.9 Interlingual Relations.

As we have said before, a specific definition for interlingual relations is not necessary here, since definition tests can be the same as the monolingual ones but using words from different languages as parameters. A problem will be to test them, since only fully bilingual reviewers could insert e.g. an English word in a Dutch test to compare it with another word.

3. Definition of the initial subset of base concepts

The selection of the core subsets of words/concepts for each language and the definition of a methodology to extend these subsets is a major concern of this task, as it is crucial to guarantee a maximal quality of the final product.

The general characteristics of the final vocabulary covered in EWN should be the following:

- 1) First of all, it must contain all concepts needed to relate other concepts in the database, namely, no semantic domain should be isolated.
- 2) There must be a high degree of overlapping across languages in the final multi-lingual WordNet. It would be senseless, within a lexical multilingual project, to contemplate unrelated areas of the lexicon for each language.
- 3) The database should contain all most frequent words in every language. An important part of the effort when building a core lexical lattice for a language must be to include those words which prove to be the more usual in current language use. For this reason one of the most important aspects in selecting the first subset of “relevant” words should be considering corpus frequency measuring.
- 4) Nevertheless, it should not be forgotten that individual wordnets are, after all, monolingual lexical nets, and therefore language-specific lexicalization patterns for common concepts have to be reflected in the database.
- 5) Specific terminology for some specialized semantic domain will be included, in order to test the ability of the EWN database to be expanded in depth. The actual terminology that will be inserted will be determined after concertation with the Interval project (which is involved in the verification of terminology databases) and taking the test-corpora of Novell in consideration.
- 6) All the vocabulary of the *Parole* project will be included to facilitate the parallel usage of both databases.

In order to optimize the relation between manual efforts and quality of the final database, it has been decided to start building the wordnets with a nuclear set of so-called *base-concepts* that should contain synsets with high positions in the hierarchy and/or high number of relations. Manual efforts will be devoted to enhance the quality of relations among base concepts and to maximize the overlapping between the set of base concepts for every language.

Base concepts have two essential properties:

- The words in that synsets are highly polysemous, they have vague and abstract meanings, complex grammatical properties, low quality definitions in dictionaries ...
- That synsets play an important role in the database; because of their position in the hierarchy, any synset is ultimately linked to them.

Although it is more difficult to establish accurate relations for such initial subsets, their quality has a major impact on the quality of the whole database, as most of the other concepts are linked to them. Thus, isolating them in a first step permits, as said above, to concentrate manual efforts where automatic procedures are less reliable.

The aim is to obtain high-quality subsets with a maximal overlapping between them, performing manual verifications and interchanging subsets between sites to balance the initial wordnets of base concepts in each language. Once this goal is achieved, it is possible to add other (more specific) areas of the vocabulary, where automatic extraction methods are more reliable.

This methodology permits to obtain balanced, high-quality individual wordnets with a relative low-cost in terms of manual effort.

The characterization of the initial subsets involves two issues:

- The selection of base concepts (ideally language independent but in fact heavily language-dependent)
- The definition of the subsets for each language covering these concepts.

As an operative criterion, the importance of a concept (that is, whether it should be a base concept or not) must be related to the ability of using this concept as an anchor to attach other concepts to its WN. As we said, this is directly related to the number of relations the concept has and to a high position of the concept in the hierarchy.

Following these criteria an initial experiment was carried on: each site was responsible for generating a first subset of base concepts and for linking it to WN1.5. Then, an evaluation of the outcomes for each WordNet and their intersections was made. We used that results to improve the background methodology and design an adequate procedure for building the subset. This methodology is presented in section 4, and the results of the procedure are described in section 5.

In the future each partner's Base Concept's subset will be extended making sure that criteria (1) through (6) above are met. Methodology will vary depending on the site. In particular:

AMS:

- 1) finalize the base concepts
- 2) extend the base concepts top-down
- 3) extend the base concepts with most-frequent corpus words which are not part of selection yet
- 4) extend the base concepts to improve overlap with other sites
- 5) add typical Dutch lexicalizations
- 6) add missing Parole words
- 7) add terminology

FUE:

- 1) finalize the base concepts
- 2) include base-concepts of other sites which miss in our subset
- 3) map (2)-concepts to WN 1.5
- 4) achieve work-in-progress automatic mappings of non-Base-Concept Spanish words to WN 1.5
- 5) link (1) through (5) subsets by hyperonymy-hyponymy relation
- 6) compare work-in-progress automatically-extracted taxonomies to those obtained by (5) in order to correct mistakes and merge both
- 7) define ways of merging (5) to automatically-extracted taxonomies
- 8) merge them
- 9) build taxonomies extending gaps in Base Concepts
- 10) add missing Parole words
- 11) add terminology

PSA:

1. Extend the base concepts top-down
2. Include words used to define base concepts which are not present in the subset.
3. Extend the subset using the most frequent words in the Italian Reference Corpus.
4. Include base concepts coming from the other sites but absent in our subset.
5. Add typical Italian lexicalizations.
6. Add missing Parole words.
7. Add terminology.
8. Finalize the base concepts.

4. Determining the subset

After some prospects and initial experiments, the following methodology for selecting the subset has been adopted. At the time of producing this deliverable, step 1 has been completed by all sites, and the others are in progress.

Here we take account of noun relations; see deliverable D006 for a similar procedure for verbs.

- STEP 1: initial set of noun senses.

For each language an initial set of word forms - grouped in synsets - has been selected, according to the criteria described in section 3. As the sources used in the different sites differ in coverage and feasibility, this criterion has been individually tuned for every specific language:

- Dutch: using the existing taxonomy as primary source, the synsets with most relations have been selected, up to a 15% of total amount of relations. Specifically, this means that all noun senses that have more than 15 relations (965 senses) have been selected. This initial selection has been adapted using the following additional criteria:

- the hyperonyms of the initial selection with 15 or less relations .
- occurring in the first three levels of the database
- tops having more than 10 related senses at any level
- occurring more than 2000 times in a 42MLN text corpus and being a top or having more than 15 relations

The final result is a list of 905 base concepts.

- Italian: We have extracted a first list of approximately 300 “base concept” items for nouns (word senses), covering nouns with 20 or more relations. We estimate that this first subset for nouns covers approx. 41% of the total amount of relations (43,686) which have been extracted from the whole set of noun definitions. The coverage of this set of base concepts is wider, i.e. 51%, when considered with respect to the size of the final subset of nouns which will be represented within the EuroWordNet database.

- Spanish: two complementary sources have been used: 1) An extended taxonomy of Spanish obtained from the monolingual Vox dictionary and 2) a manual translation of the two higher levels of WN1.5. We have also used frequency counts of words in a) the definition (and example) field of the monolingual Vox dictionary and b) a 3Mw corpus of Spanish. The criteria for selecting a Spanish noun synset have been:

- Only words appearing as translation of the two higher level WN1.5 synsets have been considered.
- For these words, if the frequency as genus word in the taxonomy was greater or equal to 5 they have been automatically selected.

- If the frequency as genus word in the taxonomy was less than 5, the word has been selected only if the number of occurrences in the dictionary was 50 or more, or the number of occurrences in the corpus was 100 or more.

The result of the selection has been 361 noun synsets and 104 verb synsets.

- STEP 2: adding missing genus concepts.

This first selection will be extended with all senses that are genus words of this selection but are not yet part of the selection themselves, because they don't fit criteria used in step 1 (this step has already been completed by AMS).

Note that this approach assumes that there are no islands in the hierarchy. All relations have to be connected in a single hierarchy to guarantee that we are not losing or missing chunks. Alternatively, this approach is applied to all the chunks of a hierarchy and the merge is taken.

- STEP 3: mapping senses to wordnet 1.5.

The augmented selection is then divided (for each language) into two different sets of synsets: set **A**, including the synsets with direct equivalence with WN1.5 synsets, and set **B**, including the remaining synsets. The procedure for splitting the selection depends, obviously, on the content of the selection. In the case of the Spanish set, all synsets belong to **A** and **B** is empty. In the other cases a manual assignment should be performed whenever possible, to build **A** (this step has been completed). Remaining synsets form **B**.

A-subsets are interchanged between sites.

- STEP 4: comparing different languages to increase overlapping.

Set **A** for each language is extended by including - as much as possible - :

- (i) Words from synsets in **B**.
- (ii) Equivalents to WN1.5 synsets that have been linked to every language but this one.
- (iii) Most frequent words in monolingual dictionaries definitions.
- (iv) Headwords in monolingual dictionaries.
- (v) defining words of a sense already included in the first selection.
- (vi) words related to the first selection by any of the defined semantic relations.
- (vii) words in bilingual dictionaries.
- (viii) words in available lists of basic words in the language (being the *Parole* lexicon a special case)

- STEP 5. Distinction between first order and higher order noun concepts.

Sheffield will study to what extent it is possible to make a distinction between first-order and higher-order noun synsets (states, events, relations, conditions) in WordNet1.5. If this distinction is persistent at all levels we can split the A-lists (see above) into both categories.

5. Base-concepts exchange

The aim of this section is comparing the overlap of the first subset of major concepts for each language. The first thing to be noticed is the differences in coverage in the three languages due to the different criteria used to select the subsets:

	synsets	words
AMS	1013	1173
FUE	362	625
PSA	336	294

As it can be seen there is a notorious difference between synsets and words in the case of FUE (1.7 words per synset). The reason is that the first criterion taken into account for Spanish has been belonging to the first two levels of WN 1.5, therefore the first task done has been to assign by hand a set of Spanish translations to WN 1.5 top and top-1 levels. The rest of the criteria used to select the final subset have been applied on this initial set. In the case of AMS and PSA the criteria used basically consisted on selecting words according to their own pre-existing taxonomies, and then mapping them to WN 1.5 synsets.

In the case of AMS and FUE the sources for comparison consist of a list of words mapped to WN 1.5 synset-identifiers. In the case of PSA, as explained in appendix I, it has not been possible to reliably map Italian words to WN 1.5 synset-identifiers -only a list of equivalent English words is supplied. This lists coincides with WN 1.5 synsets in some cases, but not always. In order to provide some data for comparison a procedure has been applied to map when possible these lists to WN 1.5 synset-identifiers. 225 of the 294 Italian words could be mapped into 392 WN 1.5 candidate synsets -without any disambiguation . Hence the reliability of this mapping is of course worse than that of the other sites (for this reason PSA figures are marked * in the table below), but this way some approximate data for Italian can be provided.

The result of intersecting pairs of subsets is the following (notice that these figures correspond to intersection of WN 1.5 synsets -not words) (see Appendix III.4 for the intersection list of Spanish and Dutch base concepts):

	AMS	FUE	PSA
AMS	1013	101	*143
FUE		362	*42
PSA			*392

The result of intersecting the three subsets is 25 synsets.

In the following, since calculations on the basis of PSA approximate data are not reliable, only AMS and FUE data are compared.

For AMS the percentage of synsets which intersect with FUE is 9.9%. For FUE with respect to AMS it is 27.9%. Obviously this difference is due to the different size of the two subsets. The low percentage of intersection is probably due to the difference in sources and approaches used by each site -as well as the different size of the subsets-. In fact, computing the intersection of the different subsets and then extending each one to cover the union can be seen as an excellent way of overtaking these differences, reaching a common subset covering a range of relevant concepts much larger and more complete than those obtained by every site on its own.

As the main goal of this merging is to extend the local subsets in order to cover a common subset of major concepts, in the case of AMS its subset will be increased by 261 FUE synsets; and in the case of FUE, its subset will be increased with 912 AMS synsets, thus resulting in a common subset of 1274 synsets. If we took the approximate data for Italian as reliable, this common subset would be increased by 207 new PSA-synsets, resulting in a final common subset of approximately 1500 synsets.

An important issue will be checking the situation of the different subsets involved within WN 1.5 taxonomy. For doing so it has been computed the number of synsets of each language in relation to WN 1.5 levels. As baseline for comparison we have chosen those synsets belonging to the nouns.top file (i.e. WN 1.5 top concepts) as well as the direct hyponyms of these (i.e. top-1 concepts).

The following distribution has been obtained. It should be noticed that up to 96% of the synsets are placed in the 4 highest levels of the hierarchy.

Depth 1 : 11 synsets
Depth 2 : 139 synsets
Depth 3 : 178 synsets
Depth 4 : 274 synsets
Depth 5 : 24 synsets
Depth 6 : 2 synsets

Total: 628

The following results have been obtained for each base-concepts subset:

Spanish Major Concepts:

Corresponding to a subset of the previous ones selected by applying the complementary criteria (see Section 4). The same ratio (96%) of synsets placed in the 4 highest level of the hierarchy is maintained. A closer view of these figures reveals that for Depth 1, 73% of the synsets are covered; for Depth 2, 70%; for Depth 3, 53%; and for Depth 4, 54%. That seems to point that the complementary criteria (number of relations and frequency) reinforces the higher levels of FUE initial choice.

Depth 1 : 8 synsets
Depth 2 : 97 synsets
Depth 3 : 94 synsets
Depth 4 : 147 synsets
Depth 5 : 15 synsets
Depth 6 : 1 synsets

Total: 362

Dutch concepts:

Due to the different way of selecting the subset the depth of the subset is clearly greater than in the previous case. Only 30% of the synsets are placed in the higher 4 levels of the hierarchy. The rest are spread over the whole hierarchy.

Depth 1 : 7 synsets
Depth 2 : 41 synsets
Depth 3 : 88 synsets
Depth 4 : 169 synsets
Depth 5 : 210 synsets
Depth 6 : 197 synsets
Depth 7 : 173 synsets
Depth 8 : 74 synsets
Depth 9 : 41 synsets
Depth 10 : 9 synsets
Depth 11 : 3 synsets
Depth 12 : 1 synsets

Total: 1013

Italian concepts:

In this case, 36 % of the synsets are placed in the first four WN 1.5 levels. The results are quite similar to those of the Dutch case.

Depth 1 : 3 synsets
Depth 2 : 14 synsets
Depth 3 : 44 synsets
Depth 4 : 81 synsets
Depth 5 : 89 synsets
Depth 6 : 67 synsets
Depth 7 : 53 synsets
Depth 8 : 23 synsets
Depth 9 : 15 synsets
Depth 10 : 2 synsets
Depth 11 : 1 synsets

Total: 392

In the following, results for the intersections between different languages subsets are offered.

Spanish-Dutch:

Most of the overlapping synsets belong to the higher level of the hierarchy, up to a 93% overlapping for this level.

Depth 1 : 6 synsets
Depth 2 : 36 synsets
Depth 3 : 24 synsets
Depth 4 : 28 synsets
Depth 5 : 6 synsets
Depth 6 : 1 synsets

Total: 101

Spanish-Italian:

Despite of the lack of confidence of these results, involving Italian data the figures are quite similar to the previous ones seeming to confirm the results.

Depth 1 : 2 synsets
Depth 2 : 14 synsets
Depth 3 : 10 synsets
Depth 4 : 13 synsets
Depth 5 : 2 synsets
Depth 6 : 1 synsets

Total: 42

Dutch-Italian:

The intersection in this case is produced at a different level of the hierarchy. Up to 40% of the synsets are placed in the first 4 levels. An interesting point to notice is that, although spread at different levels there is an important gain (from 30% to 40%) signalling that higher levels are preferred for intersection of major concepts.

Depth 1 : 2 synsets
Depth 2 : 10 synsets
Depth 3 : 13 synsets
Depth 4 : 33 synsets
Depth 5 : 33 synsets
Depth 6 : 26 synsets
Depth 7 : 14 synsets
Depth 8 : 9 synsets
Depth 9 : 2 synsets
Depth 10 : 1 synsets

Total: 143

Spanish-Dutch-Italian:

These figures seem to confirm the trend. 92% of the synsets appear in the 4 higher levels of the hierarchy.

Depth 1 : 2 synsets
Depth 2 : 10 synsets
Depth 3 : 4 synsets
Depth 4 : 7 synsets
Depth 5 : 1 synsets
Depth 6 : 1 synsets

Total: 25

A neat conclusion of these figures is that most of the common concepts are placed in the highest levels of WN 1.5. This result comes evidently from the fact that Spanish concepts are placed due to the way of selection in these levels. But intersection between Dutch and Italian seems to confirm this trend. As one of the main criteria for extending these initial subsets (see Section 3 above) consists of covering for each language the base concepts of the others, FUE and PSA should make an important effort in order to incorporate AMS concepts. The fact that most of the base concepts are high in the hierarchy will imply that most of this task will be carried out in a top-down fashion. We think that this procedure will make the task easier to perform. In the case of AMS the situation is the opposite: most of its synsets are placed in intermediate levels, so the extension should be performed both top-down and bottom-up. The increase of difficulty in this case will be lightened by the fact that the number of base-concepts to incorporate is clearly lower.

6. Experiments

6.1 Introduction

Some multilingual experiments have been carried out in order to foresee the problems we will face when building the language-specific databases (i.e. when extracting relations) and mapping them to WN1.5.

The purpose of the experiments was twofold:

- To determine the best approach for the different resources and starting points
- To explicitly define the relations and the coverage of the database

We agreed on limiting the scope of the experiments to three middle-size domains: FOOD&DRINKS, INSTRUMENT and FEELING. These fields provide a great variety of semantic relations, cover both physical and abstract concepts and were supposed to arise most of the problems we will find when trying to extract semantic relations from sources and to attach words to already existing or new synsets. Experiments were planned to cover most of the semantic relations defined in section 2 although most efforts in the initial experiment focus on the notions of synonymy and hyponymy.

The main issue along the extraction process has been to distinguish between synonymy and hyponymy/hyperonymy relations. By definition, distinguished senses cannot be synonymous. However, across dictionaries and databases we see that the sense-distinctions vary considerably. Especially in the case of sense-broadening or sense-specialization we see that the sense-distinctions may be disputable. A word such as *clean* in WordNet1.5 is differentiated with respect to the kind of dirt or pollution that is removed, whereas it has only a single meaning in the Dutch source. Such minimal variation in the denotation of meanings can be considered to be synonymous. If we widen the notion of a sense (either by grouping senses or by grouping variants) the synsets will become bigger, which may have two different side effects:

the merged senses have different hyperonyms
the merged senses have different groups of hyponyms

Multiple hyperonyms may have three different interpretations:

- the hyperonyms are synonymous and there is no problem for merging them.
- the hyperonyms are non-synonymous but compatible. They reflect different perspectives of classifying the same thing. These can be stored as conjunctive hyperonyms in the database.
- the hyperonyms are non-synonymous and incompatible. They reflect a different denotational range of things to which a word can be applied.

Only in the latter case we have to decide whether they represent different meanings or a denotation range represented by a disjunctive range of hyperonyms. Although recurrent patterns of disjunct

hyperonyms can be dealt with automatically, in other less frequent cases we have to decide this manually or by comparing the wordnets.

The merging of hyponyms can lead to problems as well. Even though the sets of hyponyms are to some extent classifiable by a unified synset, there still remain cases where one of the variants in a synset can be used to name only a subset of the hyponyms and another variant can only be used to name another subset. In the case of a merge of hyponyms we have to decide whether these denotational differences of the variants in the synset are pragmatic differences in the use of these words or genuine differences in the denotational range. In the latter case the application of the substitution test of one variant with a particular group of hyponyms leads to semantic non-sense, whereas in the former the violation of a substitution test is weaker, e.g. the intuition is “unlikely” or “unclear” rather than “no”.

There will be cases where one would prefer to keep the members of a synset (which are fully equivalent in denotation) separate from near-synonymous cases where the denotation varies a little but not sufficient to make it a separate meaning. For these cases we use the relation NEAR_SYNONYM between synsets. The implication is that the denotational variation is reflected by a separate sense, but the closeness in meaning is expressed by the explicit link.

6.2 Major conclusions

These are the main general conclusions that can be drawn from the experiment:

- I. In many cases, there are no automatic techniques available to make solid distinctions between synonymy and hyponymy. Semi-automatically enlarging the synsets leads to the merging of hyperonyms and hyponyms, which is intuitively not correct. Synsets can be enlarged either by looking for different synonymous words or by merging senses of single words.
- II. Automatic techniques currently available to build taxonomies, which include automatic word-sense disambiguation, are still not fully reliable. Nevertheless, a manual selection of the appropriate intermediate nodes on automatically derived taxonomies leads to good results.
- III. Most problems involved in both (i) and (ii) seem to occur at the higher, more abstract levels.
- IV. Most of the information extracted consists of hyperonymy-hyponymy relations. FUE, for instance, has extracted the following number of relations for the subsets of the experiment:

	Hyponymy	synonymy	antonymy	involv	part-of	set-of	made-of
food& drinks	3339	105	0	68	40	10	33
instrument	568	28	0	97	1	14	0
feeling	632	33	1	4	0	1	0

- V. automatic mapping of particular-language synsets to interlingual WN1.5 has to deal with a variety of problems, mainly due to polysemy, the linguistic coverage of the sources used in the task (namely, bilinguals and WN1.5), and the structure of WN1.5.

On the basis of these results we decided to work along two lines: (i) manual encoding of the relations for the base concepts, and (ii) (semi-)automatic procedures for the more specific levels.

In this way we make sure that we reach maximum consensus and compatibility at the highest levels of the hierarchy, where relations are most complex and the resources are least reliable, whereas we can apply the automatic techniques to the more specific levels where they are more successful. We decided to extend the experiment to the base concepts which will be divided into larger semantic clusters. The problems and solutions encountered will be discussed across the sites. For the automatic procedures we will exchange the definition patterns and the heuristics for interpreting these in terms of the explicitly defined relations. Likewise, we ensure that we are using the same ideas and interpretations across the sites.

Next sections summarize the problems that have been detected in the experiments at each site. See Appendix II for a detailed description of the individual experiments.

6.3 Result of experiment for Dutch

The general conclusion is that the Vlis (Van Dale database) relations can be taken as a starting point for deriving the EuroWordNet database. Where possible this information will be improved by:

- extracting additional information from definitions
- using the bilingual dictionaries
- comparing wordnets cross-linguistically

Special attention will be paid to relations which are unreliable or ambiguous, such as associatives of which we know that they may have lead to wrong decisions. Furthermore, some relations are missing in the Vlis database and have to be derived by systematically looking for patterns expressing it.

In general we can apply the following techniques for improving data and adding links:

- extracting definition patterns and structure for relevant relations.
- check for the badly motivated ends in Vlis relations.
- check frequent combinations of hyperonyms for synonymy, disjunct or conjunct interpretations. If disjuncts are too extreme the synsets have to be differentiated again. Since multiple hyperonyms are not allowed in VLIS in our case we have to find these multiple cases by:
 - looking for deviating and coordinated genus words in definitions
 - looking at other senses
 - checking the association and reference relations
 - enlarging the synsets with bilinguals
- look for phrases in definitions to find missing children
- look for top-concepts to find potential combinations of hyperonyms
- look at classifications in other sources to find alternative hyperonyms
- generate synsets using bilinguals to find new synonyms
- check the different meanings of the entries by looking for frequent pairs of hyperonyms that are applied to different meanings of the same word.

The extraction of information from the definitions was done in an experimental way, but have shown that many regular patterns can be found that provide additional information. This information is certainly useful provided of course that it is clear what we are looking for and what the implications are.

6.4 Result of experiment for Italian

Our starting point was to isolate the Italian data relevant for the selected taxonomies for the noun concepts represented by: food, instrument, feeling. This was done by scanning our source data to identify the set of lexical items belonging to each taxonomy; constructing word taxonomies and frequency lists for the genus terms; organizing the data into synsets; building the synset hierarchy; identifying other semantic relations apart from hyperonymy/hyponymy and synonymy. Using the methodology described in the Appendix, semantic sets of items were constructed as follows:

Feeling: approx. 700 word senses

Food: nearly 1000 word senses

Instruments: over 1000 word senses

The main problem we encountered when attempting to construct synsets from our data was that our word taxonomies are based on IS-A links between word instances. But the scope of the project is to construct a concept hierarchy. This means that the hierarchical word taxonomy must now be integrated and expanded in other directions by data on synonymy and on other kinds of semantic relations. The different relations that can be constructed have been described in Section 2 above.

This work implies a complete reconstruction of our word taxonomies as we begin to structure them into a network of relations between concepts rather than a simple hierarchy of instances. While, on the one hand, we have a simplification or flattening of our vertical taxonomies, on the other, we have a far more complex structure on other axes.

We would like to stress here, however, that while section 2 defines the global model of all the relations that can be implemented in EuroWordNet, it is obvious that not all relations can be built for each synset, nor can all relevant relations be extracted from our source data. In any case, following the WordNet philosophy, the core of the multilingual semantic network that we are building is constituted by the synsets. For this reason, the starting point for our work in the experiment was the construction of the synsets for the three selected semantic fields, and their hierarchical structuring. The source data for our synsets is a combination of data from monolingual machine dictionary synonym fields and from a synonym dictionary, integrated with data from monolingual synonymical-type definitions, and the semantic indicators in the bilingual LDB (Lexical DataBase). All the data, in particular that derived from the two latter sources, is extracted automatically but must be revised manually.

Very briefly (and simplifying), the procedure for constructing the synsets thus operates in 3 steps:

1. Explicitly tagged synonyms contained in the machine-readable dictionary entries, and synonym dictionary are grouped to form a first proposal of synset. The output is revised manually.
2. Candidate synonyms extracted from synonym-type definitions are associated with all members of the synset under construction. The output is revised manually.

3. Each candidate for the synonym set is searched in the bilingual dictionary: semantic indicators and translation equivalents are associated. The output is revised manually.

The manual revision at the end of each stage is essential: our source dictionary data is not (as dictionary data never is) totally precise in its sense distinctions; sense shifting is inevitable and cannot be left uncontrolled. In particular, the boundaries of the individual synsets are often difficult to define precisely, in particular when dealing with data on abstract concepts.

We give here an example of how using this procedure we build a synset for the concept represented by the Italian word “ansia” (anxiety).

Step 1 automatically gives us the following set of items:

ansia+++, ansietá, affanno++++, ambascia++, travaglio+, timore++, inquietudine+, pena++, apprensione, trepidazione, angoscia, /dolore+, tormento+++, afflizione++, strazio+, patimento, tristezza+, accoramento+, supplizio, sofferenza, maliconia, martirio, tortura,
(where + denotes repetition)

This example shows why revision even of explicitly tagged synonyms is necessary. At a certain point (instanced by “pena”) there is a shift in meaning from the general idea of “anxiety” to that of “sadness/suffering”. The synonym chain is thus interrupted manually at “angoscia” and, selecting the best candidates, we thus have a first synset proposal of:

ansia, ansietá, affanno, ambascia, travaglio, timore, inquietudine, apprensione, trepidazione, pena, angoscia

In Step 2, we automatically associate terms extracted from synonym type definitions for these items:

angoscia, disagio, patimento, fatica, preoccupazione, cura, struggimento, ansia, pena

of these nine items, “angoscia, preoccupazione, cura, struggimento, ansia” were accepted in the revision stage as belonging to the synset in construction, which becomes:

ansia, ansietá, affanno, ambascia, travaglio, timore, inquietudine, apprensione, trepidazione, pena, angoscia, preoccupazione, cura, struggimento.

In Step 3, we associate semantic data from our bilingual dictionary with these items and at the same time read the translations:

ansia: No semantic indicator; anxiety
 ansietá: No semantic indicator; anxiety
 affanno: No semantic indicator; anxiety, worry
 ambascia: Not in bilingual
 travaglio: No semantic indicator; anguish, distress
 timore: preoccupazione; fear
 inquietudine: No semantic indicator; anxiety, worry
 apprensione: No semantic indicator; apprehension
 trepidazione: Not in bilingual

pena: angoscia; worry, anxiety
angoscia: No semantic indicator; anguish
preoccupazione: inquietudine; anxiety
cura: No direct translation
struggimento: No semantic indicator; yearning

The only relevant semantic indicators found are “preoccupazione, angoscia, inquietudine” (words already present in the synset). In the manual revision stage, the translation equivalent data is also considered in order to help confirm the inclusion of an item in the synset under construction. It was noticeable that 7 items were translated by worry and/or anxiety:

ansia, ansietá, affanno, inquietudine, pena, preoccupazione, inquietudine

These were immediate candidates for the synset.

“apprensione”, translated by apprehension, was also considered immediately relevant. “timore”, “struggimento” and “cura” were excluded as clearly belonging to different semantic fields.

On the basis of the translation data, “travaglio” and “angoscia”, translated by anguish and/or distress, were initially considered to be potential members of another synset.

However, when an attempt was made to map our data to WN1.5, we found that “distress, worry, trouble” is a synset in the “feeling” taxonomy, with “anxiety” as its direct hyperonym.

The final synset:

ansia, ansietá, affanno, inquietudine, pena, preoccupazione, inquietudine, apprensione, travaglio, angoscia

is thus mapped to WN1.5:

distress, worry, trouble

The main problem, however, is not the construction of the synsets, or even the association of other kinds of semantic relations with the synsets, but the structuring of this data in taxonomies. The experiment has made it very clear that while it will be possible to implement automatic procedures for much of the individual synset construction, with just the need for a manual revision at the end of each step; the construction of the synset taxonomy will have to be performed to a large extent manually, and will be subject to constant revision. Again, this is particularly true for the abstract data.

As was to be expected, our three separate semantic sets behaved very differently for the construction of synsets. By far the most difficult task was represented by the set of abstract concepts belonging to the domain of “feeling”. Italian tends to be particularly rich in words to express sentiments, and to make very subtle sense distinctions. This has meant that the “feeling” synsets tend to be large but that their construction is not easy; the data has to be evaluated very carefully and there is the danger of subjective decisions. One useful test when deciding whether a candidate belongs to a given synset or not is to examine the translation equivalent; if the translation equivalent for the doubtful item is very different from the translations of the other items in the synset, then it is likely that this item does not belong to the synset under construction. On the opposite pole, the set of items for “food” contained very little synonym data; most of the synsets constructed contain just one member. Only towards the top of the taxonomy we do find some examples of synsets with multiple members. This was to be expected; most words describing items of food are very specific and there is little redundancy in this area. With respect to the phenomena of synonymy, the set of “instruments” is midway between the two extremes represented by “feelings” and “food”. As described in the Appendix, the major problem encountered in the construction of the synsets in this set was that of circularity at the top level. The differences in behaviour of the three semantic sets of data with respect to other kinds of semantic relations is discussed in the Appendix.

The clear differences between the three sample sets of data meant that they have been very useful for experimentation; we are optimistic that the strategies adopted and methodologies studied to handle the case study data will be easily extendible to cover the entire subset of word senses on which we will build our WordNet.

6.5 Result of experiment for Spanish

Since FUE's approach largely relies on both automatic mapping of Spanish words to WN1.5 synsets and automatic building of taxonomies, which will merge to give rise to the final lattice, experiments have been performed in order to both evaluate the behaviour of the automatic taxonomiser and check the accuracy of automatic mappings to WN1.5. The latter is also going to be useful to foresee the problems that any partner, regardless of the particular approach, is eventually going to face to map their wordnets to WN1.5.

Automatically-extracted taxonomies have proved to give often noisy results in the upper nodes - namely, mistakes on genus sense disambiguation. Nevertheless, they are reliable from middle-level to bottom nodes. The result is that, for instance, when building the taxonomy of 'feeling', lots of mistakes occur around this precise node, but when manually picking some nodes which are obvious 'feeling'-hyponyms (such as anger, pleasure, happiness or so), these nodes happen to be productive, and their hyponymy chains largely correct. The obvious procedure to follow seems to be, instead of tracing hyponymy chains from 'feeling' itself, to pick its appropriate hyponyms, get their taxonomies, and link them to 'feeling'.

On the other hand, mapping to WN1.5 is made via bilingual English-Spanish and Spanish-English dictionaries by using different criteria (monosemy or polysemy of the words, one or several translations in bilingual, presence of the word in one or various WN1.5 synsets...). Tests made on the automatic mappings obtained via each of the different criteria used give for the criteria which

came to be trustworthy results which range between the following values (depending on the criteria):

correct mappings: from 86 to 93%
 incorrect mappings: from 2 to 7.5%
 doubtful (mappings to a 'close' meaning): from 1.3 to 6.7%
 mappings to a more general synset: from 0.4 to 2 %
 mappings to a more specific synset: from 0.7 to 4.7 %.
 (see appendix II for details)

These are the main sources of incorrect mappings:

- (i) polysemy, either in Spanish (SP) or English (ENG), or in both, complicated by the fact that sometimes it is not reflected in bilinguals and/or WN1.5. As an example, SP 'alivio' (“get rid of a burden or load”) was mapped to WN1.5 “lightening” (“change to a light color”). The source of the error is the fact that SP-ENG bilingual translates “alivio” to “lightening”, and “lightening” is polysemous (“change to a light color” + “get rid of a burden” + ...). Moreover, there is no entry in the ENG-SP bilingual for “lightening”, and WN1.5 only bears “lightening” in its “change to light color” sense. Thus not only we got an incorrect mapping but we also have no chance of getting the correct one -nor automatic ways of detecting both mistakes. Different combinations of polysemy with gaps in both the sources and WN1.5 is the main cause for fully incorrect mappings.
- (ii) a related source of non-correct mappings is what is sometimes called polymorphysm, that is some 'weak' polysemy occurring when a word shows various slightly different meanings tightly related one to another, as what happens with 'swim' as either a sport or a plain entertainment. As in (i), this fact interacts with gaps in sources and so, giving raise in this case to mappings which can not be considered neither completely wrong nor completely correct but more or less 'around' the concept, as that of SP “baño” (translatable to “swim”, but doubtfully in its sports sense) to WN1.5 “swim” in the sense of “sport”.
- (iii) differences in granularity in the lexicalisation of concepts is the main cause of mappings to either the hypernym or an hyponym of the correct synset. When English is more granular than Spanish too specific mappings may occur -and the other way round. For instance there is no precise Spanish word for English 'hug' (which is hyponym in WN1.5 of 'embrace'); the fact that ENG-SP bilingual translates 'hug' to 'abrazo' (which is a more general concept) causes the mapping 'abrazo' to WN1.5 'hug', when the correct mapping would be 'abrazo' to WN1.5 'embrace'. As told, the reverse case causes mappings to too general concepts -e.g. SP 'colada' to ENG 'washing', being a 'colada' something more specific than a 'washing'.
- (iv) last, some errors are ascribable to formal properties of MRDs, namely different types of either typing mistakes or non-systematicity in the coding of information.

Appendix I: Extraction of base concepts

I.1 Extraction of Dutch subset

The Van Dale database Vlis is different from the traditional MRDs in that particular semantic relations between senses are already explicitly coded. We therefore can make use of some other criteria than the other sites. The following relations are in the Vlis database:

Semantic Relations in the Vlis database

	Noun Senses
ABBREVIATION	63
ANTONYM	181
ASSOCIATIVE	1763
CAUSATIVE	0
HYPERONYM	49815
INCHOATIVE	0
PARTITIVE	576
REFERENCE	715
SYNONYM	9901
WOMAN	95
PREFERENCE	2361
FORM_VARIANT	4
SYMBOL	1

The database does not form a *closed* wordnet in which all relations are unified in a single tree or a small set of tops. In fact, the Vlis database contains 1429 ends for the nouns. The reason for making a sense a top or end in the hierarchy is often not well-motivated (the database is still in development). In many cases a sense is an end because the meaning was too complex (e.g. many higher-order-nouns referring to states, conditions, events) or information was of an encyclopaedic nature (names of places, people, etc..). We therefore did not just take the Vlis tops as a starting point for selecting the base concept but we considered the following criteria to be important in order of appearance:

- number of relations
- position in hierarchy
- Vlis top senses
- frequency

In total 905 noun senses have been selected. In the next sections we explain how this selection was achieved.

First criterion: number of relations

Inspection of the Vlis database lead to the conclusion that the major criterion for being a base concept seems to be the number of relations a sense has.¹ To get a sensible and manageable first selection in a consistent way, we first selected the senses with the most-frequent-relations making up around 15% of the total amount of relations. For this purpose we constructed a table with the cumulative-percentage of all relations, ranking the senses per number of relations. This is shown in the next table (Table I.1).

Table I.1

Nouns				
Number of Relations	Number of Senses	Number of Links	Percentage of total number of links	Cumulative percentage of total number of links
0	0	0	0,00%	0,00%
1	53998	53998	37,28%	37,28%
2	10420	20840	14,39%	51,67%
3	3646	10938	7,55%	59,22%
4	1675	6700	4,63%	63,85%
5	1107	5535	3,82%	67,67%
6	739	4434	3,06%	70,73%
7	467	3269	2,26%	72,99%
8	405	3240	2,24%	75,22%
10	247	2470	1,71%	76,93%
11	199	2189	1,51%	78,44%
12	177	2124	1,47%	79,91%
13	131	1703	1,18%	81,08%
14	117	1638	1,13%	82,21%
15	110	1650	1,14%	83,35%
16	82	1312	0,91%	84,26%
17	75	1275	0,88%	85,14%
18	68	1224	0,85%	85,98%
19	69	1311	0,91%	86,89%
20	48	960	0,66%	87,55%
.....
514	1	514	0,35%	100,00%
28605	74280	144842	100,00%	0,00%
1to9	72457			
>9	1823	2,45%		
>12	1200	1,62%		
1to15	73438			
>15	965	1,30%		

All noun senses with more than 15 relations resulted in 965 senses (1,3% of all senses) whose relations cover circa 16% of all noun relations.

¹ At this stage all relations have been taken into account to constitute the importance of a concept.

We realised that senses of this first selection may have hyperonyms higher up in the hierarchy that do not have that many relations, but still can be base concepts. Therefore we generated all the hyperonym chains up from the senses of this first selection and extended this first selection with these senses. So our second selection contained all nouns with more than 15 relations united with all the hyperonyms of these senses up in the taxonomy. The results are in table I.2:

table I.2

Noun senses

> 15 relations	hyperonyms	union
965	470	1165

Second criterion: position in hierarchy

This last selection still contained many senses that we would not consider as base concepts. Another important criterion seemed to be at which level in the hierarchy a sense is positioned. The higher a sense is located in the hierarchy the more likely it is that it is a base concept. In order to determine the level of a sense we generated a top down taxonomy starting from the top senses for nouns. As downward relations, relations that imply a different degree of specificity, we used the hyponymy and meronymy relations for the noun taxonomy. In general, relations that do not imply a different degree of specificity like synonymy and antonymy could be disregarded for the levels of the hierarchy. However in the Vlis database we had to consider these as well because most of these relations, except synonymy, were allowed to have downward relations themselves. Therefore we extended each level with these sideward relations, i.e. for nouns: abbreviation, antonymy, associative, preference, form-variant, symbol and woman. We disregarded the reference relation, because it functions as a morphological link between parts of speech. The number of senses at each level downwards and sideways are in table I.3, whereas table I.4 contains the number of senses per level, where the levels are united.

Total number of Vlis noun senses = 72704

Number of Vlis noun tops = 1429

table I.3

Level	number of senses downwards	number of senses sideways	number of senses
0	1429	732	2161
1	5585	572	6051
2	11.511	1061	12382
3	16.110	1334	17.198

table I.4

Union of noun levels

levels	number of senses	percentage of all senses
0 and 1	8154	11,2%
0, 1 and 2	19661	27,0%
0,1, 2 and 3	32295	44,4%

These levels were needed to filter out the non-base concepts of our second selection. Again we used the percentage of the total amount of senses to draw a line. The number of noun senses down to level 2 is 27 % of the total amount of noun senses. We decided to take the intersection of these levels with our first selection of most-frequently related senses.

Intersections

Nouns with more than 15 relations and their hyperonyms = 1165
intersected with noun top down taxonomy

table I.5

levels	intersection (number of senses)	percentage of the first selection (1165 senses)
level 0	148	12,7%
down to level 1	432	37,1%
down to level 2	706	60,6%
down to level 3	921	79,0%

The cut-off point was determined by inspecting the lists of included and excluded senses per level for too specific senses.

Relevant Vlis top senses

So far we generated base concepts first in a bottom up manner by taking the most-frequently related concepts, extended with their hyperonyms. This selection includes only a fragment of all the ends/tops in the Vlis database. This is shown in the previous table I.5, where the intersection of the most-frequently-related senses and the levels is given. Level 0 here represents the level of the Vlis ends or tops. The following tables show the overlap of the Vlis ends with the selection so far:

table I.6

Not selected Vlis ends

	total of Vlis ends	Intersection with selection so far	Not selected Vlis ends
nouns	1429	148	1281

Because the Vlis taxonomy is still in development many of these ends are not real top senses, and therefore irrelevant for our subset. To filter out the relevant ones we used the top down hierarchy and counted all the children down from a so-called top sense (at any level). We draw the line at 10. This resulted in 114 Vlis ends with 10 or more senses related to it (any depth). From these 114, 35

are new nouns which have not been included in the previous selection. The union of these results in the final selection of 741 noun senses.

Finally: addition of frequency.

At AMS we only have frequency information at the word level. This appeared to be less reliable as a criterion for a basic semantic concept. Therefore, we only included frequent words when they also satisfy one of the other criteria: many-relations or being a Vlis end. This resulted in the addition of another 164 noun senses. The final subset of base concepts then consists of the union of this list with the previous result: **905 noun-senses**.

Translation of the Dutch base concepts

The above base concepts have been translated to equivalents in WordNet1.5. For each sense one or more translations have been given. The following problematic situations have been encountered when translating the base concepts:

- 1) a VlisSense is broader and corresponds with several WN1.5 senses. In that case multiple senses are selected. This can either mean that the VlisSense is a hyperonym or that the WN sense-distinctions are too fine-grained. Criteria for deciding that the senses are synonymous are:
 - if the selected senses have the same hyperonyms or other relations
 - if the selected senses have the same conceptualization (especially in those cases where the denotation is less important or can vary).

Examples:

keuze:2 = choice:1 0091731-n {the making of a choice}
 choice:2 3929489-n {what is chosen}

- 2) multiple VlisSenses can be linked to a single WN sense (the opposite situation). In this case each VlisSense is linked to the same WN sense. This either suggests that the senses are synonymous (should belong to the same synset) or the WN sense is a hyperonym. Later on, when the wordnets are compared it will turn out that non-synonymous senses are linked to the same ILI-record and a similar choice has to be made as in the previous case.

Examples:

hout: 1 {the substance trees consists of}
 = 09057553-n wood 4 {the substance and any kind of these substances}
 houtsoort 1 {the kind of substance}
 = 09057553-n wood 4 {the substance and any kind of these substances}

versiering: 1 {what is used to decorate}
 = 02029323-n decoration 2
 versiersel: 1 {what is used to decorate}
 = 02029323-n decoration 2 {what is used to decorate}

In the case of “wood” we see that the meaning type of wood is lexicalized in Dutch by a separate word. Since any word can refer to some entity or the type of this entity this is not expressed in the

English meaning of “wood”. In the case of “decoration” we see that both words in Dutch have exactly the same meaning and thus have to be joined in the same synset as full equivalents.

- 3) a VlisSense represents an intermediate level between a WN sense and its hyperonym. The WN sense is too specific and the hyperonym is too general.

Example:

toon: 1	(too narrow)	04471153-n	tone	7
	(to broad)	04731716-n	sound	5
stem: 2	(too narrow)	03477899-n	voice	1
	(too narrow)	04599795-n	voice	5
	(to broad)	04731716-n	sound	5

In this case both the more specific and the more general WN synset have been linked to the Dutch sense.

- 4) a VlisSense is equivalent or has a strong relation to a synset with another POS:

arme: 1 Noun	XPOS_BE_IN_STATE	01550906-a	poor 5 Adjective (poor people)
verval: 1Noun	XPOS_NEAR_SYNONYM	00122638-v	decline 5 Verb
vastzitten: 1 Verb	XPOS_NEAR_SYNONYM	00131922-a	stuck 1 Adjective

- 5) A good clue for equivalence are the sets of hyponyms, meronyms etc.. linked to a VlisSense-WNSense pair. However, there can be two problems here:
- the downward relations match but the definition and/or upward relations (the meanings) do not.
 - the meanings and upward match very well but the hyponyms, meronyms etc.. differ considerably.

Here the definitions and upward relations (the true meaning of the sense and not its classifying role) should be taken as a starting point. The difference in classification can be due to other factors. Hyponyms can be classified in different ways (e.g. by function or constitution) or at different levels. The latter is very likely because WN uses many technical intermediate levels. These differences in classifying will follow from the comparison of the wordnets.

In some cases the suggested translations seems inappropriate. In that case it helps to browse up and down the taxonomy to find alternatives. More specifically it helps to see whether the translation of a hyponym is linked to a hyperonym in WN which is more appropriate. Translation-equivalence can thus also be generated in a bottom-up fashion.

Another general observation is that the meanings in WordNet1.5 are often differentiated whereas only a single or less meanings are given in the Dutch source. For example, the Vlis only gives one

best linked to WN1.5 “factory, mill, manufacturing plant, manufacturing” or to its direct hyperonym “plant, works”. See also *malattia 1* and *colpo 1*. This suggests to us that these senses of WN1.5 should perhaps be merged in the Interlingual Index (ILI). Otherwise, we presume that the best strategy is to map directly to the higher synset in the WN hierarchy.

b) similarly, it often occurs that a single Italian item can match equally well to more than one WordNet synset. For example, we have “*oggetto 2*” which maps to both “aim, object, objective, target” and also “purpose, intent, intention, aim”. As these two synsets both belong to the same taxonomy (which terminates in “psychological feature” passing via “goal, end”), it appears reasonable again to propose a merging between the WN1.5 items for our Interlingua. However, a proposal of this type is probably not feasible when our Italian item matches to WN1.5 entries which belong to different taxonomies, e.g. we have “*stato 4*”, translated by the bilingual LDB into “state”, which has currently been mapped to WN1.5 “state, province, territory”; WN 1.5 “country, state, land, nation”; WN1.5 “state, nation, country, land, commonwealth, res pubblica, body politic”. In this case, the first WN1.5 entry is in the “location” taxonomy, whereas the other two belong to “group, grouping”. This suggests that probably the Italian entry should be revised and perhaps split into two senses.

c) indeed, frequently a single sense in the Italian data is already clearly split by our bilingual LDB into more than one sense; in such cases again we will probably be creating separate senses in the Italian WordNet. An example of this is “*macchina 1*” which in fact encapsulate the very different senses of “machine”, “engine” and “car”.

d) another, less frequent problem, is when we can find no WordNet equivalent sense, e.g. the Italian “*elenco*” is naturally translated as “list”; however, the only WN1.5 entry under list is “list, listing (a database ..)”. This is clearly a limited sense of list and not that implied by “*elenco*”. See also “*locale 2*” and “*scritto 2*”. In these cases, presumably, we must add a new sense to our Interlingua.

At the moment, where our data suggests that more than one of our base concept items can be matched to the same (or very close to) WordNet synset, we grouped them together and indicated this fact. See for example, “*atto 1*” and “*azione 1*” which both appear to map to WN1.5 action 1. A decision will be taken as to whether these sense should be merged later on. Otherwise, the list has been ordered for decreasing frequency value of genus term (these values must be taken as approximate for now - at times we have items which appear frequently as genus terms in more than one sense; in these cases the frequency value given is, currently, the overall value, and has not yet been broken down sense by sense, e.g. *segno 1* (WN1.5 mark, sign) and *segno 2* (WN1.5 symbol 1) have each been assigned the overall frequency value of 78 occurrences.

(ii) Construction of Base Concept Synsets

The preliminary synsets are being constructed by associating with the candidate item data extracted automatically from the monolingual LDB synonym field and from a synonym dictionary, integrated with data from monolingual synonymical-type definitions, and the semantic indicators in the bilingual LDB. The two latter sources are extracted automatically but must be revised manually.

Thus, for example, the definition in our LDB for “*congegno, 2*”, which has been matched to WN1.5 “mechanism 4”, is “*ordigno, meccanismo, dispositivo*”, while in the bilingual LDB we have two semantic indicators assigned to this basic sense of *congegno* “*dispositivo (device), meccanismo (mechanism)*” The proposed synset is thus currently formed by: “*congegno, ordigno, meccanismo,*

dispositivo”. This synset will probably be merged with that currently attached to another of our base concept lexical items “dispositivo, 1” (again matched to WN1.5 mechanism 4) which is currently assigned to synset proposal “dispositivo, meccanismo” on the basis of data extracted from the bilingual LDB.

However, in this context, we will also have to consider, for instance, the synset proposed for “apparecchio, 2” Here, the synset is currently constructed on the basis of definition data: “apparecchio, congegno, dispositivo”. “Apparecchio” has currently been matched to WN1.5 “device 1”, which is the hyperonym of WN1.5 “mechanism 4”. We will have to decide whether to keep the “congegno, ordigno, meccanismo, dispositivo” and the “apparecchio, congegno, dispositivo” synsets separate or merge them into one concept and match them to one entry of ILI. All decisions of this type will only be possible when we have had the chance to examine our taxonomical data in more depth. One point that is emerging very clearly is that as we proceed with the construction of the synset taxonomy, careful manual revision will be necessary and the positions in the synsets in the taxonomies will vary.

As can be seen, our work on the top concept construction (and creation of relevant synsets) is still in an early stage. The problems encountered have for the moment been left open; more definitive decisions will be taken when we have had more opportunity to analyse our data in further depth and compare our first results with those of our partners in the project. In this first list we thus present separate lexical items - not yet mapped into synsets. The work on the synset construction is thus still at a preliminary stage.

1.3 Intersection between Spanish and Dutch subset

Relation of WN1.5 synsets present both in the Spanish and in the Dutch subset.

Overall number: 101 synsets

synset first word form

00002728-n life_form
00003711-n cell
00004865-n person
00008030-n animal
00008894-n plant
00009469-n object
00010368-n substance
00011263-n food
00011607-n artifact
00012356-n article
00012878-n cognition
00013299-n motivation
00013522-n feeling
00014558-n shape
00014882-n time
00015245-n space
00015437-n state
00016649-n act
00017008-n group
00017394-n possession
00017586-n attribute
00018966-n measure
00019295-n phenomenon
00228990-n activity
00507790-n production
00736689-n young
00740781-n microorganism
01633257-n larva
01958400-n thing
01958716-n thing
01959683-n whole
01965302-n fabric
02003723-n drug
02009476-n instrumentality
02028879-n opening
02029323-n decoration
02032220-n plaything
02034531-n structure

02329807-n commodity
02486678-n surface
02855539-n part
03056705-n sheet
03282629-n trait
03338771-n quality
03607347-n body
03841132-n ability
03944302-n information
04111788-n attitude
04139704-n message
04155501-n language
04313427-n message
04326789-n show
04425761-n signal
04599795-n voice
04788545-n desire
04792478-n pleasure
04827440-n temper
05074818-n beverage
05116306-n world
05116476-n people
05120211-n collection
05432072-n line
05443777-n point
05449837-n region
05450515-n region
05637686-n rock
05638634-n remains
05650477-n part
05715416-n body_of_water
05720524-n land
05846273-n expert
05856677-n worker
05918609-n acquaintance
06102108-n friend
06163124-n relative
06276483-n power
06464347-n natural_phenomenon
06465491-n consequence
06467144-n luck
08239006-n process
08434357-n room
08440487-n connection
08453309-n kinship
08483587-n figure
08484352-n line

08520394-n condition
08522741-n situation
08523567-n relationship
08523811-n relationship
08530753-n nature
08531278-n degree
08550427-n disorder
08711637-n union
08781633-n material
08783090-n mixture
08805286-n chemical_element
08907331-n compound
08918157-n element
08936946-n fuel
09028514-n poison
09065837-n time_period

Appendix II: Experiments on specific subsets

II.1 Detailed description of Dutch experiment

II.1.1. Introduction

The following noun synsets from WN1.5 were chosen for the experiment:

WN1.5 synsets	Dutch equivalents in Vlis
FEELING <feeling> (the psychological feature of experiencing affective and emotional states)	gevoel_3
DRINK <beverage, drink, potable> (any liquid suitable for drinking)	drank_2
FOOD <food, nutrient> (any substance that can be metabolised by an organism to give energy and build tissue)	voedsel_1
INSTRUMENT <instrument, tool> (the means whereby something is accomplished; “my greed was the instrument of my destruction” or “science has given us new tools to fight disease”)	instrument_1

The Dutch experiment first of all consists of a critical assessment of the data stored in the Van Dale database Vlis with respect to the data requirements of EuroWordNet and an investigation of the possibilities to improve this information using the information extractable from the definitions and the bilingual dictionaries.

The Van Dale database is different from a traditional Machine Readable Dictionary in that some semantic relations between word senses are already explicitly coded. How the different VLIS relations for the Dutch equivalents are distributed in the Vlis database at their next level is shown in the following tables (for an explanation of the relations see below):

	SYN	HYPO	ANTON	ASSOC	REF	PREF
gevoel3	2	49	-	2	-	-
drank2	1	37	-	-	-	-
voedsel1	7	47	-	4	-	-
instrument1	-	37	-	1	-	-

	HYPER	PARTI	WOMAN
gevoel3	1	-	-
drank2	2 levels up	-	-
voedsel1	-	-	-
instrument1	2 levels up	-	-

In the following sections these relations will be evaluated.

II.1.1.1 The sense-distinctions

The nouns that were chosen for the experiment consists of a higher order noun: feeling and first order nouns: drink, food and instrument. We first looked at the senses that were distinguished in Vlis for these words.

GEVOEL (feeling)

sense1 het vermogen om stoffelijke voorwerpen of druk- en pijngewaarwordingen te herkennen, te voelen
(the ability to recognise, to feel material objects or pressure and pain sensations)

sense 2 wat men voelt, de lichamelijke gewaarwording van het voelen
(what one feels, the physical sensation of feeling)

sense 3 = 4 het vermogen van de mens om iets innerlijk gewaar te worden en te waarden
(the ability of the human being to perceive something mentally)

sense 4 = 3 wat men innerlijk gewaarwordt
(what one perceives mentally)

Sense 1 of gevoel (feeling) is synonym with 'tactile sense', which meaning is not within the scope of this experiment. Sense 2 and 4 are disjunct meanings, divided in 'physical' and 'mental' feeling. Sense 3 denotes 'the ability to feel': the feeling for or sense of something or someone. For some reason the senses 3 and 4 are swapped when the taxonomy was build, from now on we will talk of 3 with the definition of 4 and vice versa.

Both sense 2 and 3 are linked to 'gewaarwording' (sensation); the physical feeling (sense 2) is a synonym and the mental feeling sense 3 is a hyponym of 'gewaarwording' (sensation). This is an inconsistency, according to the definitions both senses should be a hyponym of 'gewaarwording'

(sensation), differentiated in physical and mental. Another possibility is to take up these two senses in one and link it as a synonym to 'gewaarwording' (sensation).

The question arises what the criterion could be for disambiguation here. When 'gevoel' (feeling) and 'gewaarwording' (sensation) are put in a synset their hyponyms will be the union of mental and physical feelings. The list of these hyponyms will contain for example the items: 'medelijden' (compassion), which is a mental feeling; 'jeuk' (itch(ing)) which is a physical sensation and 'pijn' (pain) which can be both. If 'gevoel' is divided in a mental and physical sense, then entries like 'pijn' (pain) that have both senses will have to be a hyponym of both the mental and physical sense. See also hyponymy.

drink

The first sense of drank denotes 'medicine'. The second one is the 'liquid to drink'. The third and last one is 'alcoholic drink'. The last one seems to be a hyponym of the second sense, which also is the case in WN1.5

food

Voedsel_1 (food) has only one sense, which seems correct.

instrument

Sense 1 of instrument has the following definition (translated literally):

more or less complex or delicate tool or apparatus for some sort of work by hand, business or scientific research. Sense 2 has a one-word definition which is 'hulpmiddel' (aid, means). Sense 3 denotes the instrument of music.

Sense 1 and 2 of instrument refer to a different property of the same object. E.g. a 'vergrootglas' (loupe) is an instrument to see small things bigger, which can be used by someone who is partially sighted as a 'hulpmiddel' (aid). There are instruments/ tools that we would immediately denote as 'hulpmiddel' (aid) like 'krukken' (crutches) and others like 'stethoscope' which we would define by its use.

Therefore we believe sense 1 and 2 to be near synonyms. See also hyperonymy and hyponymy.

Conclusion

It is important to have good sense distinction for these major concepts; it will make life easier for building the rest of the taxonomy.

II.1.1.2 Synonymy

The noun gevoel_3 (feeling) has two synonyms in Vlis: 'gevoelen_1' and 'sentiment_1'.

The definitions of gevoel_3 and its synonym "gevoelen_1" are more or less similar:

gevoel_3	: "wat men innerlijk gewaarwordt"	(what one senses internally)
gevoelen_1	: "wat men innerlijk voelt"	(what one feels internally)

The verbs used in these definitions are synonymous. So the definitions may be considered equal. To have similar definitions seems to be a correct criterion for being synonyms. Provided that they consist of multiple words.

The second synonym 'sentiment_1' is defined with one word, namely:

"gevoel" (feeling). It is not clear to what sense of 'gevoel' (feeling) it refers. In the bilingual VDL 'sentiment' is translated as 'sentiment' with the variants: 'feeling' and 'emotion'.

One-word definitions do not clearly determine synonymy.

The noun 'drank_2' (drink) has one synonym: the noun 'drinken_2' (drink). Comparing the definitions results in:

'drank_2': "vloeistof die men opdrinkt" (liquid that one drinks)

'drinken_2': "vloeistof om te drinken" (liquid for drinking)

The first definition just states the fact that the liquid is potable, the second definition expresses the fact that the liquid is meant for drinking. This information denotes the relation: involvement (patient-role).

The noun 'voedsel_1'(food) has 7 synonyms, they are listed below with the corresponding definitions, countability and style:

	<i>def.</i>	<i>tr.</i>	<i>count</i>	<i>style</i>
eten 1	voedsel	(food)	no	-
etenswaar 1	voedsel voor mensen	(food for people)	no	-
eetwaar 1	--	(--)	-	-
kost 3	wat men eet	(what one eats)	no	-
spijs 1	bereid voedsel	(prepared food)	yes	archaic
voeding 2	voedsel	(food)	no	-
vreten 2	eten	(eten)	no	vulgar

First of all the definitions of these senses are very short and therefore not very clarifying. But when the definitions do contain some distinguishing information like 'for people' or 'prepared' it has apparently not affected the synonymy relation. However the senses express many differences in various aspects.

'Eten' (food) is a general mass term for anything that is edible and what is eaten by creatures that eat. 'Etenswaar' (comestibles) is uncountable and also expresses what is eaten, but it is restricted to food for people. It is used in the context of what we buy in the shop and we would talk of the storage life/tenability of 'etenswaar' (comestibles). The third sense is just a form variant of sense 2. Sense 4 'kost_3' (fare) is a mass word and expresses food in general; it is used however only with adjectives like: daily, plain, simple, light and heavy. It is not what you buy, but it denotes the prepared food on the table. Sense 5 'spijs_1' (food), a countable word, also denotes the food that is prepared, but it is slightly archaic. It is mostly used for food in chic restaurants or it refers to biblical scenes or food eaten by gods. Therefore it looks more like a hyponym of voedsel (food). Sense 5 'voeding_2' (nourishment), non-countable, denotes food that we need to nourish ourselves. We talk of 'eenzijdige voeding is slecht voor je gezondheid' (unbalanced nourishment is bad for your health). The last sense vreten_2 has the definition 'eten' (food) which is correct but its style is rather vulgar.

Synonymy seems to have been taken in a very broad sense, but it is not clear in what sense.

Conclusion

We need to define synonymy considering aspects like:

- definitions
- syntactic differences
- contextual behaviour
- countability
- register.

II.1.1.3 Hyperonymy

All nouns of the subset are positioned relatively high in the Vlis hierarchy.

The hyperonym of *gevoel_3* is *gewaarwording_1* (sensation). This noun itself has no more hyperonyms but is related with a reference relation to the verb '*gewaarworden_1*' (to sense), see also section 1.7 reference.

We already mentioned that the noun *drank_2* has only one hyperonym: *vloeistof_1* (liquid) while its definition and the definition of its synonym imply also a relation with the verb 'to drink' (involvement: agent-role). Generation of these links for nouns in Dutch could be done by parsing the definitions and look for patterns like: 'om te V' (for V-ing)

Instrument_1 has a hyperonym, namely the noun *voorwerp_1* (object; def. = case, object that one perceives, treats etc.). We suggested that *instrument_2* (aid, means) is a near synonym of *instrument_1*; this implies that the hyperonym of *instrument_2* has to be assigned too. This hyperonym is '*middel_2*' (means) with the definition: what is used for achieving a certain goal. This seems an appropriate hyperonym for both senses.

Conclusion

The hyperonyms in Vlis can be taken as a starting point. They have to be checked for consistency by checking the taxonomy structure. Not all the information from the definitions in the monolingual source has been used in Vlis to apply proper hyperonyms. Other methods like extraction from the definitions or comparison with WN1.5 may help to find these. Furthermore we need reliable tests that can be used for verification of the relation.

II.1.1.4. Hyponymy

Most hyponyms of gevoel_3 contain head words in the definition like: gevoel van/ voor/ te/ dat (feeling/sense of/ to/ that); (gemoeds)toestand (state (of mind)); stemming (mood); het X(adj.) zijn (being X-adjective). Other examples are compounds with the noun 'gevoel' (feeling) on the right hand side. Some examples with the definitions literally translated.

HYPONYM	definition
bewondering_1 (admiration)	gevoel van bewonderen (sense/feeling of admiration)
schuldgevoel_1 (sense of guilt)	het gevoel schuld te hebben (the feeling to have guilt)
angst (fear)	gevoel dat er onheil of gevaar dreigt (feeling that danger is coming)
opwinding_2 (excitement/ tension)	toestand van overspanning (state of overstraining)
emotie_1 (emotion)	geëmotioneerde gemoedsgesteldheid (emotional state of mind)
blijdschap_1 (joy, gladness)	aangename stemming (pleasant mood)
vrolijkheid_1 (cherrfulness)	het vrolijk zijn (being cheerfull)

These definitions imply that feeling, state of mind and mood belong to one synset. The situation in Vlis for the relevant entries is illustrated below:

toestand (state)		
HYPO		
gesteldheid (constitution)		
HYPO		
<stemming1;	(mood)	<ASSOCIATIVE> gevoel_3 (feeling)
SYN bui2;	(mood)	
SYN humeur1	(mood)	
SYN lui1;	(mood/humour)	
SYN geestegesteldheid2;	(mental state)	
SYN gemoedsgesteldheid2;	(emotional state)	
SYN gemoedstoestand2>	(state of mind)	

This taxonomy shows a large synset of synonyms but somehow gevoel_3 (feeling) is not interpreted as such but linked by an associative relation. We believe that 'mental state' is not a proper synonym but should be higher up in the hierarchy; 'mood' is a hyponym of 'state of mind'. 'Mood' and 'feeling' are strongly related but are not synonyms:

It is a mood therefore it is a feeling?

It is a feeling therefore it is a mood?

We think 'mood' is a complex of feelings, therefore we suggest that 'feeling' is a co-hyponym of 'mood' with either a meronymy or a near-synonym relation with 'mood'.

Evidence for near-synonymy is the fact that their hyponyms do not form a natural class altogether. We realise that the difference between mood and feeling is not always sharply distinguished. The new situation would be:

toestand (state)

|HYPO

gesteldheid (constitution)

|HYPO

geestesgesteldheid (mental state)

|HYPO

<gemoedsgesteldheid (emotional state); SYN gemoedstoestand (state of mind)>

|HYPO

<stemming; bui; etc.> (mood/ humour)

?NEAR-SYNONYM

gevoel (feeling)

|HYPO

?MERONYM

|HYPO

rotbui (rotten mood)

haat (hate)

hoerastemming (state of jubilation)

schuldgevoel

(sense of guilt)

The hyponyms of drank_2 contain many items that seem to be related too high in the taxonomy. This is partly due to the fact that it is not always clear for drinks which hyperonym is to be applied. This problem already occurred in the Acquilex project. In Vlis it seems that lexicographers have been careful with hyponymy relations based on the right element of a compound or the head words in the definitions. 'aniseed milk' has not been related to 'milk' and 'liquorice water' has not been related to 'water'.

HYPONYMS definition (literally translated)

of drank_2 (drink)

anijsmelk: 1 melk, gekookt met anijszaad en suiker

(aniseed milk) (milk, cooked with aniseed and sugar)

dropwater: 1 water waarin drop is opgelost, soms met wat citroen

(liquorice water) (water in which liquorice is solved, sometimes with a little lemon)

Hyponyms of voedsel_1 (food) also consist of a large range of various items. A number of hyponyms are compounds ending on one of the synonyms: -kost, -eten, spijs, etc. for example: burgerkost_1 (plane fare). Also definitions that contain as head word one of the synonyms often lead to a hyponymy relation. This causes wrong hyponyms; 'borstvoeding' (breast-feeding) should

be a hyponym of voeding_1 (the feeding). The problem is disambiguation of the words used in the definitions.

Most hyponyms of instrument_1 use in the definitions the head word 'instrument'. Some examples that do not are (definitions are translated literally):

bril_1 (glasses)

(means for improving the sight or for protecting the eyes, consisting of two glasses in a frame placed in front of the eyes.

contactlens_1 (contact lens)

lens that one, by means of replacement for glasses, wears directly on the eyes.

gastroscoop_1 (gastroscope)

apparatus with which one can look inside the stomach

The first hyponym 'glasses' is according to the definition a hyponym of instrument_2 (means). The second example looks like a hyponym of 'lens' or a hyponym of instrument_2. As we suggested this sense to be a near synonym of sense 1 the hyponyms have to be applied correctly. The third example has the head word 'apparatus' in the definition, which fact leads to the question why it has not been assigned a hyponym of 'apparaat' (apparatus), see also associative relation.

Conclusion

The hyponymy relation in Vlis can be used as a starting point but has to be checked and can be improved. For this we may use the definitions; compound rules and comparison with WN1.5.

II.1.1.5 Antonymy

The antonymy relation has been assigned 181 for nouns. Examples for this relation within the subset of the experiment are hard to find. Under the taxonomy of gevoel_3 (feeling) we found:

minderwaardigheidsgevoel	<->	meerderwaardigheidsgevoel
(feeling of inferiority)	<->	(feeling of superiority)

One would expect many more occurrences, e.g.:

liefde	<->	haat
(love)		(hate)
wantrouwen	<->	vertrouwen
(distrust)		(trust)

In Dutch, antonymy is often related to morphological properties, examples of prefixes with negation are 'on-', 'wan-'. But also the definitions contain elements of negation that illustrate antonymy, e.g. 'wantrouwen_1' has the definition: afwezigheid van vertrouwen (absence of trust).

Conclusion

Antonymy has not been assigned consistently in Vlis. Morphological properties, negative elements in the definitions and comparison with WN1.5 may be used to improve the relation. We also need a good definition and test for verification.

II.1.1.6 Associative

In VLIS this relation is used when the hyperonym relation was not satisfying or when there seemed to be a strong association with another entry which could not be expressed otherwise. This relation can be cross-part-of-speech. Within the Vlis database the relation has been assigned 1763 times to a noun. Within the scope of the experiment we found occurrences of an associative relation for *gevoel_3* (feeling) with *stemming_1* (mood), which we converted to a near-synonym and/or meronymy relation, discussed at hyponymy. It also has an associative with *bewustzijn_1* (consciousness/ awareness). The definition of this entry is: *vermogen tot besef, tot weten en erkennen van het bestaan van zichzelf en van de wereld* (ability of awareness, knowing and acknowledging the existence of oneself and the world). It is not clear to what other relation this associative relation could be converted. All relevant entries are linked properly, so this associative can be deleted.

food

Voedsel_1	has 4 associatives:	hyperonyms
baksel_1	(baking/ batch)	voorwerp_1 (object)
fruit_1	(fruit)	teelt_2 (crop)
groente_1	(vegetables)	teelt_2 (crop)
gerecht_1	(dish/ course)	none

All four examples have poor links in the hierarchy and the associative relation is the only way by which these entries are linked to 'food', which is implied by all definitions except for batch; fruit: edible fruits; vegetables: vegetable food for humans as plant or food; dish: food served in one course. Associative could be converted in hyponymy relation.

Instrument_1 has an associative with *apparaat_1* (apparatus); both are related to *voorwerp_1* (object). The definition of *apparaat_1* (apparatus) is literally translated: more or less composed tool. This is part of the *instrument_1* definition. There seems to be a near-synonym relation but they have different hyponyms. In Dutch we use *apparaat* (apparatus) for tools that are electrical like: *televisie* (television); *elektrische tandenborstel* (electric toothbrush). While non-electric tools like microscope are called instruments. This difference is not expressed in the definitions and would be a too fine-grained difference. Other words that seem to constitute the concept in Dutch are: *toestel* (appliance), *machine* (machine), *installatie* (installation/ machinery), *werktuig* (tool). More information from bilinguals or WN1.5 is needed to get a hold of the concept.

CONCLUSION

In most cases the associative relation can be converted to another relation which helps to improve the taxonomy. Sometimes the relation is superfluous which seems to be the case when the items involved in the relation are already linked properly.

II.1.1.7 Reference

In VLIS this relation denotes morphological relationships between all categories and it has been assigned 2746 times to nouns.

We found a reference relation for the noun *gewaarwording* (sensation) with the verb *gewaarworden* (to sense). In EWN we would express this as an XPOS-has synonym relation between N-V.

Down in the taxonomy many reference relations are between the nouns and their corresponding adjectives, which we would express in EWN as an XPOS-has synonym relation between N-A.

afkeer (aversion)	reference	→	afkerig (averse to)	XPOS-has synonym N-A
bangheid (fear)	reference	→	bang (frightened)	XPOS-has synonym N-A

Another example of a noun to verb reference is:

vrees (fear)	reference	→	vrezen (to fear)	XPOS-has synonym N-V
-----------------	-----------	---	---------------------	----------------------

Other examples are found under drank2:

dorstlessen (thirst-quencher)	reference	→	dorstlessend (thirst-quenching)
sap (juice)	reference	→	sappig (juicy)

This link with the adjective, which is a sort of property of the noun, cannot be expressed semantically in EuroWordNet.

II.1.1.8 Partitive

The partitive relation in Vlis is an extra relation that stands for an undifferentiated meronymy relation. We found the following examples for part/ whole and member/ collection meronymy:

Partitives for drank_2 (drink)

bier_1 partitive (beer)	→	hop 1 (hops)	part/ whole
----------------------------	---	-----------------	-------------

definition of bier_1 is: alcoholhoudende drank uit hop en mout door gisting bereid.

(alcoholic drink prepared by fermentation from hops and malt)

First of all it is strange that 'mout' (malt) has not been assigned as a partitive to bier (beer). The specification 'part/whole' of this meronymy relation can be extracted from the definition of the meronym, i.e. 'bier' (beer) . Although the definition uses the verb 'prepared' we do not believe this relation should be specified as a 'made of' meronymy.

Partitives for instrument_1:

beeldbuis_1	partive	→	televisie_1	part/whole
(screen)			(television)	

Definition of beeldbuis1 (screen): elektronenstraalbuis waarop het beeld in een televisietoestel gevormd wordt. (cathode ray tube on which the picture is formed in a television)

The specification that it is a part/ whole relation can this time be extracted from the holonym.

deegroller_1	partitive	→	keukengerei_1	member/collection
(rolling pin)			(kitchen ware)	

The meronymy: member/ group relation cannot be extracted from the definitions of these items.

Conclusion

The partitive relation in Vlis can be used as starting point for EWN. The relation has however not been applied consistently and it is not differentiated enough. The Vlis partitive relation should be converted in a more general one like meronymy. To specify and improve this relation we need methods to extract information from the definitions and other sources.

II.1.1.9 Woman

Another relation that occurs in VLIS but not in WN1.5 is the woman relation, which is a link between female and male form of nouns. It has been applied 95 times.

Some examples (outside the scope of the experiment):

actrice_1	→	acteur_1
(actrice)		(actor)
advocaat_1	→	advocate_1
(lawyer)		(lawyer)
alcoholicus_1	→	alcoholica_1
(alcoholist)		(alcoholic)

Conclusion

This relation will not be used in EWN, but the information could be converted as a co-hyponym with a label stating which the form is used for males or females. In many cases the form that is

used for males is also the form used by default. In that case the ‘woman’ link should be converted into a hyponym.

II.1.1.11 Preference

The preference relation is used to refer to other entries which are preferred and where the relevant information should be found. For nouns it occurs 4476 times.

Haag_1 (Hague)	preference	-->	Den Haag_1 (The Hague)
nulzes-nummer (written out form)	preference	-->	06_nummer_1 (06-number)

Conclusion

This relation will also not be used in EWN, but the information could be converted as a synset variant with a label stating it is the preference.

II.1.1.12 Symbol

The relation is between the symbol and the entry it symbolises. It has been applied 202 times for nouns. This relation will also not be used in EWN, but the information could be converted as a synset variant with a label stating it is the symbol. Examples are:

Celsius	symbol → C
Bulgarije (Bulgaria)	symbol → BG
watt	symbol → W

II.1.1.13 Abbreviation

The relation is between the abbreviation and its full form. It has been applied 63 times for nouns. Again this relation will not be used in EWN, but the information could be converted as a synset variant with a label stating it is the abbreviation.

Examples are:

host: 1	hostcomputer	
ktv: 1	kleurentelevisie	(colour television)
micro: 1	microcomputer	
pc: 1	huiscomputer	(home computer)
tel.: 1	telefoon	(telephone)
tv: 1	televisie	(television)

II.1.1.14 Form-variant

This relation can be converted as a synset variant; it has been applied for 8 nouns.

II.1.2. Extracting information

We experimented on generating (missing) relations and improving less reliable relations in Vlis by using information from various sources and by different methods.

Synonymy

To extract synonyms we can use the following resources

- monolinguals: all words with exactly the same definition, explicit synonymy relations
- bilinguals: all words which have the same translation(s)
- WordNet1.5: translation of variants in the synsets

In the case of the monolingual resources we looked at words with the same definitions and for explicit synonymy links encoded in another Dutch dictionary (van Sterkenburg and Pijenburg, 1984). The list below shows the result for the entry 'gevoel':

Synonym candidates:		Vlis
(gevoel 22920 (1 2) N)	(feeling)	X
(sensatie 65918 (1 2) N)	(sensation)	-
(emotie 17571 (1) N)	(emotion)	-
(aandoening 98 (3) N)	(feelings)	-
(ontroering 49489 (1) N)	(emotion, poignancy)	-
(alteratie 2768 (4) N)	(commotion)	-
(consternatie 12219 (1) N)	(consternation)	-
(opschudding 50845 (2) N)	(commotion)	-
(ontsteltenis 49562 (1 2) N)	(consternation)	-
(gevoelen 22921 (1) N)	(feeling)	synonym
(gevoelen 22922 (2) V)	(to feel)	-
(ervaren 18021 (1) V)	(to experience)	-
(gewaarworden 22964 (3) V)	(to sense/feel)	-
Search completed!!!!!!		

Conclusion

Of this list only *gevoelen*(1) was assigned in Vlis as a synonym. The items of the list are all related closely to the concept of *gevoel_3* (feeling), but are not true synonyms. In Vlis most of them are related as hyponyms or hyperonyms of *gevoel_3* (feeling) We need to verify this by good synonymy tests.

For extracting synonyms from translations we use an Acquilex-tool which generates all possible English translations of a set of Dutch source words in the bilingual Dutch-English dictionary (Martin and Tops, 1986) and next looks up these English words in the bilingual English-Dutch dictionary to see what Dutch translations are generated. In those cases that one of the Dutch source words was found among the translations of the English word the translations have been proposed as possible candidate for translation. These cases form a so-called translation-circle: we return back to one of the source words via the bilingual dictionaries. If the Dutch translations of the English translation contains new words these are offered as potential synonyms of the Dutch source words.

Each of the translation-circles consists of the English word followed by a list of the translations and translation phrases.

In the following lists we find the result of this process starting with “gevoel” (feeling):

Translation Circles:

(feel (gevoel greep))
 (feel (aanleg feeling flair gevoel))
 (touch (tastzin gevoel techniek gevoel_voor_de_bal))
 (sensation (gevoel aandoening sensatie gewaarwording zintuiglijke_gewaarwording))
 (sense (gevoel vaag_gevoel zin besef instinctief_besef begrip))
 (feeling (gevoel tastzin))
 (feeling (begrip gevoel medeleven))
 (feeling (gevoel besef sensatie gewaarwording))
 (feeling (emotie gevoel sentiment gevoeligheden gevoelens))
 (feeling (idee intuïtie indruk gevoel))
 (feeling (aanleg feeling begrip gevoel zin))

Conclusion

This list contains no more additional candidates for synonymy. Most of these candidates express the sense ‘feeling for’ which is not the correct sense. ‘Sentiment’ was already applied as a synonym in Vlis. A test has to verify this relation.

A manual scrolling through the bilinguals resulted in a few additional correlating words, namely:

gemoedsaandoening (emotion, feeling)
 gemoedsbeweging: (emotion, feeling)

We also applied these methods to the near-synonym of gevoel (feeling), namely stemming_1 (mood).

The search for synonyms resulted in the following (comparison with Vlis):

		Vlis
(stemming 70262 (1) N)	(mood)	central
(humeur 28833 (1 2 3) N)	(mood)	syn
(luim 40679 (1) N)	(mood)	syn
(sfeer 66093 (1) N)	(atmosphere)	-
(bui 10132 (3) N)	(mood)	syn

Conclusion

‘Sfeer’ (atmosphere) might be a near-synonym of ‘stemming’ (mood).

Those that were not found but were applied by Vlis are:

SYN geestesgesteldheid2; (mental state)
 SYN gemoedsgesteldheid2; (emotional state)
 SYN gemoedstoestand2 (state of mind)

The last one ‘gemoedstoestand2’ was generated by automatic translation via bilinguals. Furthermore two additional ones were found:

sentiment (sentiment)
 geest (spirit)

The tests will have to determine how these words are related to the concept of mood and feeling. Furthermore a side effect of these search options is finding possible new synset items on the English side, e.g.: ‘tune’

drank_2 (drink)

A search for synonyms via the monolingual or the bilingual gave no result. Translation of other items of the WN1.5 synset (beverage, potable) did not yield any more synonyms. So the Dutch synset would be as in Vlis <drank_2 (central), drinken_2>.

Synonyms found for voedsel_1 (food) compared to those in Vlis:

voedsel 80833 (1) N	(food)	Vlis	
(spijs 68607 (1) N)	(food)	X	
(eten 18101 (1) N)	(food)	syn	
(eetwaar 17139 (1) N)	(comestibles)	syn	
(levensmiddelen 39016 (1) N)	(provisions))	--	(New Word)

New words found by translation:

(voeder 80814 (1) N) (fodder)
 (voer 80863 (1) N) (feed)

Conclusion

These words are synonym to ‘voedsel_1’ (food) but are restricted to food for animals.

Synonyms for instrument

Explicitly defined as synonyms in the monolingual are:

(instrument 30574 (1 2) N) (instrument)

(apparaat 3416 (1) N)	(apparatus)
(toestel 74305 (1) N)	(apparatus, appliance)
(machine 41125 (1) N)	(machine)
(installatie 30516 (1 3) N)	(installation, machinery)
(gereedschap 22366 (1 2) N)	(tools, equipment)
(gerei 22378 (1) N)	(gear, utensils, kit)
(gerief 22404 (2) N)	(comfort)
(benodigdheden 6267 (1) N)	(needed equipment)
(werktuig 84471 (1) N)	(piece of equipment)

New words found by translation:

(gebruiksvoorwerp 21047 (1) N)	(implement, appliance, utensil)
(comfort 11724 (1) N)	(comfort)
(mechanisme 42023 (2) N)	(mechanism)
(inrichting 30356 (5) N)	(construction, installation)
(tuig 75551 (1) N)	(gear, equipment)
(uitmonstering 76326 (3) N)	(equipment, outfit, kit)
(uitrusting 76392 (1) N)	(equipment)
(outillage 51452 (1) N)	(equipment)
(apparatuur 3417 (1) N)	(apparatus, machinery)
(materieel 41905 (1) N)	(material, equipment)
(machinerie 41137 (1) N)	(machinery)
(systeem 72247 (10) N)	(system)
(mechaniek 42018 (1) N)	(mechanism)

We did not find any new synonyms (on top of the previous results) by translating variants of the equivalent WordNet1.5 synsets.

Hyperonymy

We used the following methods to find hyperonyms:

- monolinguals: by head words in the definitions
- WN1.5: hyperonyms

Vlis:

<drank_2, drinken_2> (drink)

→ hyperonym = <vloeistof_1> (liquid)

→ hyperonym = <stof_4, materie_1> (substance, matter)

WordNet 1.5

Sense 2

<beverage, drink, potable>

→ food, nutrient

→ substance, matter --

→ object, inanimate object, physical object

```

→ entity
  → liquid
    → fluid
      → substance, matter
        → object, inanimate object, physical object
          → entity

```

In WN1.5 the synset of 'drink 'sense 2 has two hyperonyms: <food, nutrient> and <liquid>. The Dutch synset <drank_2, draken_2> has only one hyperonym: vloeistof_1 (liquid) which has as a hyperonym: stof_4 (substance). Which has a synonym materie_1 (matter). So in fact it is the synset <stof_4, materie_1> (substance, matter). This is almost analogue to the hyperonym path of <liquid> in WN1.5 except for the fact that WN1.5 has a sub level <fluid> and two more levels above <substance, matter>, which are <object, inanimate object, physical object> and <entity>.

Considering the definition in Vlis for vloeistof_1 : (transl.: each substance in state of aggregation) it is correct that its hyperonym is <stof_4, materie_1>. (substance, matter). The translation of <fluid> generates the word fluidum_2, which is in Vlis a co-hyponym of vloeistof_1 (liquid). The definition of fluidum_2 is: (tr.: gas or liquid). Does this make it an intermediate level between 'substance' and 'liquid' ?

The two highest levels in WN1.5 cannot be generated from the definitions of the Dutch entries:

```

stof_4:      materie, substantie (matter, substance)
materie_1:   stof (substance)

```

We will have to decide whether we generate top-levels based on:

- the WN1.5 topconcepts
- the definitions
- results of the experiment in which we compare top concepts from each site.

In WN1.5 it is chosen to have a second noun hyperonym; <food, nutrient>. This hyperonym results in a translation voedsel_1 (food, def. = al that can serve as nutrition, nourishment). This seems a correct second hyperonym.

Hyponymy

For hyponymy relations we used:

- keywords and genus words in the definitions
- compound rules
- comparison with WordNet 1.5.

A search for the keywords [gevoel van] (feeling/sense of) resulted in 24 items.

Some examples are :

```

afgunst_1    (envy)
schaamte_1   (shame)

```

Both examples are synonym to words that are hyponyms of gevoel (feeling).

An attempt to find missing hyponyms by compounds ending on -gevoel resulted in the 11 items. The next list shows some examples with the information of what relations they have in Vlis:

Vlis	
angstgevoel_1 (sense of fear)	synonym of angst_1 (fear)
evenwichtsgevoel_1 (sense of balance)	hyponym of vermogen_3 (ability)
haatgevoel_1 (hatred)_	preference haatgevoelens_1 (feelings of hate)

A final method for finding new synonyms is by translating the hyponyms of the WordNet1.5 synsets. An example is the word 'pain' which is a hyponym of the mental feeling. The Dutch equivalent is the word 'pijn'. In Vlis there are two senses; sense 1 is defined as "lichamelijk lijden" (physical suffering), sense 2 as "verdriet" (grief). Sense 1 is a hyponym of 'gewaarwording1' (sensation) which is related too high. Sense 2 is synonym of 'verdriet_1' (grief) which is a hyponym of gevoel3 (feeling) which is correct.

Antonymy

In order to find antonymy relations we used morphological properties, negative elements in the definitions and comparison with WN1.5

gevoel

Prefixes expressing negation are -wan, -on, etc.

A search for the prefix '-wan' and '-on' results in many occurrences, which have to be checked manually:

An interesting example is 'wanbegrip' (misconception) which is in Vlis a synonym of onbegrip (lack of understanding), however no antonymy relation has been assigned. The definitions of these nouns are:

'wanbegrip'; verkeerd begrip (wrong understanding)

'onbegrip' = geen begrip (no understanding)

Even the definition of 'onbegrip' (no understanding) implies an antonymy relation with 'begrip' (understanding). The synonym relation is dubious here.

Meronymy relation

Definition patterns of

part/whole:

deel van (part of)	accuplaat (accumulator plate)	def. = deel van accu (part of accumulator)
onderdeel van	zaagvoet	def. = deel van zaagmachine

(subpart of)	(saw foot)	(part of chain saw)
van X (, Y en Z) (of X (, Y and Z)	stamppot (mish mash)	def. = 'van groente, vlees en aardappelen' (of vegetables, meat and potatoes)
bestaand uit/ bevat of/ contains)	spiegeloptiek	def = optisch stelsel dat uit spiegels bestaat (consists of bolle spiegels bevat. (optic system that consists of mirrors or contains mirrors)

member/group

lid van (member of)	Ex. clublid lid van een club (member of a club)
---------------------	--

Other relations

The other relations are hardly present in the Vlis database and we therefore fully have to rely on the information extractable from the definitions. The selected subsets for the experiment are however not always representative for frequent definition patterns.

1) Higher-order nouns:

First order nouns will not have XPOS synonymy and/ or XPOS hyp(er)onymy relations, therefore we did not look for corresponding definition patterns in the taxonomies of 'instrument'/ 'food' and 'drink'. A search for definition patterns in the taxonomy of the higher-order noun gevoel (feeling) implying these relations gave no results.

2) ROLE relations:

2.1 ROLE_INSTRUMENT

N = X	gebruikt om te Vinfinitive	(used to Vinf)
Ex.	schuurpapier (sandpaper) (gebruikt om mee te schuren) (used for sanding with)	
	gebruikt voor het Vinfinitive	(used for Vinf)
Ex.	seinvlag def. = (signal flag) gebruikt voor het doorgeven van informatie (used for transmitting information)	
	gebruikt bij het Vinfinitive	(used at/ with Vinf)
Relations:	N	ROLE_INSTRUMENT
	'schuurpapier'	Vinf 'schuren'

	(sandpaper)		(to sand)
N = X	voor het Vinfinitive		(X for the Vinf)
N = X	bestemd voor het Vinfinitive		(X intended for the Vinf)
Ex.	stuwdam (dam) bestemd voor het stuwen van water (intended for the damming of water)		
Relations:	N	ROLE_STRUMENT	Vinfinitive
	'stuwdam'		'stuwen'
	(dam)		(to dam)
N = X + CLAUSE, VP	in which a "met PP" occurs with Y as head		
Ex.:	hengelaar = iemand die met een hengel vist (angler)		(someone who fishes with a fishing rod)
Relations:	Y	ROLE_INSTRUMENT	V
	'hengel'		'hengelen'
	(fishing rod)		(angle)

2.2 ROLE_AGENT

N = X + CLAUSE, VP (active form), where X is the first argument of an action
 bakker = iem. die brood bakt
 (baker) = (someone who bakes bread)

Relations:	N	ROLE_AGENT	V
	'bakker'		'bakken'
	(baker)		(to bake)

2.2 ROLE_PATIENT

N = X + CLAUSE, VP (active form), + NP (direct object) with Y as head
 aardappelrooier = iemand die aardappels rooit
 (potato digger) (someone who digs potatoes)

Relations:	Y	ROLE_PATIENT	V
	'aardappels'		'aardappelrooien'

3) INVOLVEMENTS

The patterns for INVOLVEMENTs are parallel to the definitions of the nouns except for the fact that the verbs always have the form of an infinitive VP:

V = VPinfinitive + with PP, X as head
 knippen = met een schaar afsnijden)
 (to cut) (to cut of with sissors)
 Relations: V INVOLVED_INSTRUMENT X

V = action VPinfinitive + ARG1 (X), ARG2 (Y)
 slachten = <van slachters> <dieren> doden voor consumptie
 (to slaughter) = (killing by slaughters of animals for consumption)
 Relations: V INVOLVED_AGENT X
 Relations: V INVOLVED_PATIENT Y

V = process VPinfinitive + ARG1 (X)
 sneeuwen = <van sneeuw> vallen
 (to snow) (<of snow> falling)
 Relations: V INVOLVED_PATIENT X

4) CHANGE_TO_STATE or CAUSES

V = higher-order-noun veroorzaken/ tot gevolg hebben (cause/ resulting in)
 V = tot higher-order-noun leiden (lead to)

Ex. gewenning = het wennen aan het gebruik van medicijnen, wat kan
 (habituation) leiden tot verslaving
 (which can lead to addiction)

Relations:
 V CHANGE_TO_STATE higher-order-noun (if noun is non-dynamic)
 V CAUSES higher-order-noun (if noun is dynamic)

II.2 Detailed description of Italian experiment

The experiment has been performed along the following basic steps that we will describe in depth below:

1. Scanning our data to draw up a list of lexical entries belonging to the three semantic fields: FEELING, FOOD and INSTRUMENTS
2. Constructing frequency lists of genus terms
3. Structuring our taxonomies
4. Reorganizing the data in synsets

II.2.1 Scanning the data

- a) Delimitation of a semantic field: sentimento (feeling)

The analysis of the Italian data regarding the field represented by the English term FEELING has been more difficult than for FOOD and INSTRUMENTS, because the answer to the question: "what is a feeling?" and the decision as to which Italian lexical items belong to this field is not simple.

We thus started from the definition of sentimento (feeling) that we found in an Italian dictionary:

sentimento: (1) ogni stato affettivo della coscienza; ogni moto soggettivo dell'animo che dia una particolare tonalità affettiva, di gioia o di dolore, alle nostre sensazioni, rappresentazioni, idee (2) l'affettività in generale; la sensibilità, la finezza di sentire (3) modo di sentire, di valutare soggettivamente qlco (4) la consapevolezza dei propri atti (feeling: (1) every affective state of the conscience; every subjective impulse of the soul that gives a particular affective tone, of joy or pain, to our sensations, representations, ideas (2) emotivity in general; sensibility, delicacy of feeling (3) the way to feel, to value something (4) awareness of one's own actions)

However our task is not simply to consider how a concept and its hyperonyms and synonyms are defined in the dictionary to draw the semantic borders for that concept: we have to study the mapping between the field occupied by the Italian concept we are considering and that filled by the corresponding English concept. Therefore, we must consider the definition of FEELING in WN 1.5:

the psychological feature of experiencing affective and emotional states
 actively exploring with the hands
 manual-genital stimulation
 the perception of tactual or proprioceptive or gut sensations
 a vague idea in which some confidence is placed
 the general atmosphere of a place or situation

Only the first meaning of the English word feeling corresponds totally with Italian word sentimento: feeling has a larger meaning than sentimento; it does not regard only a purely emotional and sentimental environment but comprises the complete sphere of physical and tactual sensations and perceptions. In the English language, time seems to have preserved a stronger semantic link between the verb to feel and its noun feeling, both of which describe emotional and physical processes and states. Whereas, in Italian, although the verb sentire covers the whole range of emotional and physical sensations, its related noun sentimento is only used for affective and emotive states. If we want to achieve a consistent mapping between the English wordnet and the Italian one we must follow one of these strategies:

- a) to widen the field of meanings on the Italian side with some other meanings, such as sensazione (sensation), percezione (perception) and atmosfera (atmosphere)
- b) to limit the semantic field of feeling to its first meaning reported by WordNet 1.5.

Following this second strategy we have extracted, so far, approximately 700 words denoting types of feelings from our sources.

During the analysis of this data there have been a few doubtful cases when facing nouns related to both feelings/condition and feelings/quality. As a general rule we have excluded nouns of quality, so we have included, for example, ira (anger) and we have discarded irascibilità (irascibility).

Generally we asked the question whether an emotional state or condition was lasting or momentary, given that many feelings have the characteristic of being sudden and compelling but not lasting. However, in a few cases the choice is not simple.

b) Semantic field: cibo (food)

The work of searching for words denoting food in our source data has been less difficult, because it concerns concrete nouns and the borderlines of the semantic field are easier to draw. Food is translated by our bilingual LDB as cibo and the Italian dictionary defines cibo in the following way:

cibo (1) tutto ci che serve da nutrimento SIN. alimento 2 l'atto del mangiare, pasto (food (1) everything that serves for nutrition syn. aliment 2 act of eating, meal)

We considered only the first meaning of the word and we decided to include in the semantic field of FOOD both foodstuffs and their raw materials, so both pane (bread) and farina (flour) are considered as food.

We have extracted nearly 1000 words for this semantic field.

c) Semantic field: strumenti (instruments)

We have considered the following three semantic features as characterizing different kinds of instruments and therefore different sub-taxonomies:

- complexity (simple tools or apparatus / set of instruments)
- use modality (manual / not manual tool)
- telos / purpose (work, art, sport, travel, war etc.)

Starting from a few key words for each of these features such as, for example, strumento (instrument), apparecchio (apparatus, set of instruments), macchina (machine), veicolo (vehicle), we were able to select more than 1000 items.

II.2.2 Frequency lists

a) Semantic field: feeling

For FEELING it is possible to draw up the following list of the most frequent genus terms, in decreasing order, with $n \geq 5$ (frequency, italian genus terms, translations):

- 62 sentimento (feeling)
- 47 stato (state)
- 30 desiderio (desire)
- 22 dolore (pain)
- 15 amore (love), cosa (thing)
- 14 turbamento (perturbation)
- 13 agitazione (agitation), disposizione (inclination)
- 11 sensazione (sensation)

- 10 senso (sense)
 9 moto (motion), passione (passion)
 8 avversione (adversion), ci che (what), mancanza (lack)
 7 coraggio (courage), forza (strength)
 6 afflizione (affliction), ansia (anxiety), disagio (uneasiness), discordia (discord), fascino (fascination), impeto (impetus), impulso (impulse), paura (fear)
 5 affanno (worry), attenzione (attention), condizione (condition), entusiasmo (enthusiasm), essere (to be), volontà (will)

b) Semantic field: food

- 74 pesce (fish)
 67 frutto (fruit)
 61 pianta (plant)
 42 dolce (sweet, cake)
 38 bevanda (drink)
 35 cibo (food)
 34 vino (wine)
 27 carne (meat)
 26 pasta (dough, pasta)
 20 formaggio (cheese)
 18 liquido (liquid), liquore (liqueur), prodotto (product)
 16 fungo (mushroom), vivanda (victuals)
 15 pane (bread), sostanza (substance)
 13 pietanza (dish)
 12 arbusto (shrub)
 10 animale (animal), salume (various kinds of cold meats and salami), taglio (cut of meat)
 9 pasto (meal), uccello (bird)
 8 acquavite (brandy), gallo (cock), minestra (soup)
 7 crostaceo (crustacean), mammifero (mammiferous), mollusco (mollusc), panino (sandwich), zucchero (sugar)
 6 acqua (water), chicco (grain), maiale (pig),
 5 alimento (aliment), bibita (soft drink), biscotto (biscuit), caffè (coffee), condimento (flavouring), gallina (chicken), genere (kind), latte (milk), uva (grapes)

c) Semantic field: instruments

- 290 strumento (tool, implement; mus. instrument)
 196 apparecchio (apparatus; set; instrument; device, appliance)
 114 arnese (tool, implement; utensil)
 92 dispositivo (device, contrivance; gear)
 65 macchina (machine; engine)
 61 attrezzo (tool, implement; pl. sport. appliances)

46	congegno (device, contrivance; mechanism)
35	veicolo (vehicle)
33	mezzo (means)
30	utensile (tool, implement, utensil)
20	arma (weapon, arm)
15	meccanismo (mechanism)
11	apparecchiatura (apparatus)
9	autoveicolo (motor vehicle)
8	automobile (car)

So far we have extracted:

- Three taxonomies for feeling:
stato (state), sentimento (feeling, within the taxonomy of "state"), sensazione (sensation)
- Four taxonomies for food:
alimento (aliment), cibo (food), liquido (liquid), pasta (pasta, dough)
- Four taxonomies for instrument:
strumento, arnese, attrezzo, utensile.

II.2.3 Structuring the taxonomies

In order to illustrate the task of taxonomies structuring we discuss the examples of feeling and instruments because they raised many more problems than did the food taxonomy.

a) instruments

The main difficulty we have to deal with, starting from our automatically created nouns taxonomies, is the inconsistency derived from the definitions themselves, in such a way that it is impossible to manage this data without an heavy manual intervention. If we consider, for example, the field of instrument/tool we find:

- circularity in the definitions of the top concepts, so we have to find a suitable criterion to decide the right hyperonym relations within the taxonomy / ontology;
- different hyperonyms for elements of the same synsets;
- different genus terms used for identical types of objects;

In Italian the most general and comprehensive word for the English "instrument/tool" is strumento and this is actually the most frequent genus term in this field used to define 290 items.

Unfortunately two basic problems arise, derived from the inconsistency of the definitions: strumento has as hyperonym the word arnese which is not perceived as more general; strumento has only two word senses, the first gathering all its concrete meanings, the second the figurative and the extensive ones.

The first point gives rise to a problem of circularity because arnese has as synonym utensile and hyperonyms attrezzo or strumento; while, in its turn, attrezzo has hyperonyms arnese (w.s.1) and strumento (w. s. 2), and finally utensile has as hyperonym arnese.

This circularity gives rise (and can be considered as a proof) to a first synset: strumento, arnese, attrezzo, utensile. But, if we consider the more general use of strumento and also its possibility of being employed in figurative and extensive senses we should place this word on a higher level

(compared with the other three) within the taxonomy. As a matter of fact in Italian we can define as strumento nearly all types of tools, but the same is not true for arnese or attrezzo or utensile:

la zappa é uno strumento (the hoe is a strumento)

il computer é uno strumento (the computer is a strumento)

la zappa é un arnese / un attrezzo (the hoe is a arnese / attrezzo)

* il computer é un arnese / un attrezzo (the computer is a arnese / attrezzo)

From the second point it follows that very different types of instruments are listed under this genus; so we find mixed simple manual instruments, scientific measuring instruments and musical instruments. In this and in similar cases we would need a more fine grained distinction with respect to our sources, which will give rise to a greater number of sub-taxonomies, based on other features which are found in the "differentia" part of the definitions. Furthermore this contrasts with the fact that other less general terms have more word senses and are better specified: for example attrezzo has three word senses, the third of which defines an homogeneous type of attrezzi: the appliances. The last problem to be observed is concerned with the different genres used to define strongly related objects such as, for example, pieces of cutlery. Examining the data we find forchetta (fork) under arnese, but neither coltello (knife) nor cucchiaino (spoon) which can be found under strumento and posata respectively. For cases like this, we need to correct the incoherency by using the appropriate level in the taxonomy for all the related words, i.e. the lowest appropriate level (in this case posata, which in turn will point to utensile and thus to strumento).

b) feeling

At a first glance the taxonomy appears heavily crowded with terms which do not seem to belong to this subset, or which have particular usage as far as register or idioms are concerned. To simplify this taxonomy and make it more coherent we carefully and precisely controlled the group of words occurring under the feeling taxonomy. As result of this check, we were able to exclude many items on the basis of both register codes and word formation type. So we decided to rule out all the words having the following register code:

fam. (family), volg. (vulgar), lett. (literary), reg. (regional),
scherz. (playful), ant. (ancient), popol. (popular)

and as semantic code:

fig. (figurative), est. (extended)

The type of word formation we considered as an indication of words not belonging to the set of FEELING were the derivatives -ismo and the scientific compound ending in -filia and -fobia. So words such as fanatismo, altruismo, campanilismo, patriottismo, bibliofilia zoofilia, gallofobia were discarded. After this, a few words still remained within the taxonomy which could not be excluded by means of these criteria. A few of them such as debole, "avere il debole del vino" ("weak point" in "wine is my weak point ") or demone, "avere il demone del gioco" ("demon" in "the demon of gambling") enter in idiomatic expressions or locutions (which are hard to handle automatically) and others need some linguistic tests in order to be considered and selected according to not

completely subjective criteria. We decided to test them as objects of the verbs sentire and provare (to feel).

This test helped us to easily discard a few of these items such as amoretto, flirt, idillio, idealitá etc., indeed in Italian we cannot say:

* sentire	un amoretto	
	un flirt	un idillio
	una idealitá	una irreligione
	una religione	bontá
	egoismo	

In conclusion, by means of these criteria, we reduced the taxonomy of feelings, which appears more strongly related and ready to be transferred from the taxonomic structure to the synonymic one.

II.2.4. Reorganization of taxonomies in synsets

Sketches of our taxonomies show how the organization of meanings coming from our sources is mainly vertical; we must now create a semantic structure that will still be guided by a hyponymy relation (@), but that can be organized in groups of synonyms (synsets). This is a radical change: taxonomies must be deeply reorganized around horizontal, synonymical cells that will cut the semantic tree trasversally.

We have studied the construction of synsets for the branch of timore (dread) in the FEELING taxonomy. The situation from which we started is the following:

STATO (state)

SENTIMENTO (feeling)	TREMARELLA (shivers)	PAURA (fear)
	ORRORE (horror)	
	TERRORE (terror)	
	TIMORE (dread)	
		ALLARME (alarm)
		PANICO (panic)
		PAURA (fear)
		cacarella
		FIFA
		FOBIA
		spaghetto
		TEMA (fear)

Many items can stay together in the same synset. The following are some definition analyses:

allarme 1 ordine gridato ai soldati ...2 grido o segnale di pericolo 3 (fig.)timore, ansia
alarm 1 order to soldiers 2 danger signal 3 (fig.) fear, anxiety

paura 1 stato d'animo costituito da inquietudine e grave turbamento, che si prova al pensiero o alla presenza di un pericolo 2 timore, preoccupazione

fear 1 mood of restlessness and perturbation in front of danger 2 dread, worry

orrore 1 sentimento di forte paura SIN. terrore, raccapriccio 2 cosa che desta tale sentimento

error 1 feeling of strong fear syn. terror, dread 2 a thing that causes this feeling

terrore 1 sentimento di forte sgomento, intensa paura SIN. spavento, orrore 2 cosa o persona che sgomenta

terror 1 feeling of strong dismay and fear sin. fright, horror 2 person or thing causing dismay

timore 1 sentimento di ansia, di sgomento, perlopi=F9 non grave, che si prova davanti ad un pericolo o ad un danno vero o supposto 2 sentimento di affetto

dread 1 feeling of anxiety, dismay, not serious, felt when faced with real or supposed danger 2 feeling of affection.

tema1 (lett) timore

fear 1 dread

There is synonymy that is directly labeled in the dictionary with the tag SIN. (terrore 1 SIN. spavento, orrore), and there is a type of synonymy that comes from the definition pattern (paura 2 timore, preoccupazione), when the definition is constituted by single words, without a "differentia" part.

We indicate, for convenience, the first kind of synonymy with SIN., the second with SYN.
We obtained the following synonymic configurations:

synset 1:

terrore 0_1 sin. spavento, orrore (terror 0_1 sin. fright, horror)
orrore 0_1 sin. terrore, raccapriccio (horror 0_1 sin. terror. dread)

synset 2:

allarme 0_3 syn. timore, ansia (alarm 0_3 syn. dread, anxiety)
paura 0_2 syn. timore, preoccupazione (fear 0_2 dread, worry)
preoccupazione 0_1 syn. apprensione, ansia sin. inquietudine (worry 0_1 syn. concern, anxiety sin. uneasiness)
tema 1_0 syn timore (fear 1_0 syn dread)

apprensione sin. inquietudine, timore, ansia (uneasiness sin. agitation, dread, anxiety)
Synset 1 {orrore 0_1, raccapriccio 0_0, terrore 0_1, spavento 0_1}
Synset 2 {allarme 0_3, ansia 0_1, apprensione 0_0, inquietudine 0_1, paura 0_2, preoccupazione 0_1, timore 0_0, tema 0_1}

When we group together lexical items in synsets because they are synonyms, they often have, in the sources, different hyperonyms: this is our most serious problem.

Explanatory examples:

-orrore 0_1 (horror 0_1) and terrore 0_1(terror 0_1) are linked to the same hyperonym sentimento (feeling), whereas raccapriccio 0_0 (dread) comes from turbamento (perturbation).
turbamento (perturbation) is not in the taxonomy of stato (state) with sentimento (feeling): it is in a branch apart and comes directly from the verb turbarsi (to get upset) through a verb to noun (v2n) relation.
-preoccupazione (worry) does not have as hyperonym sentimento (feeling), but it is linked to preoccuparsi (to be worried) by a v2n relation.

We will have to use, as superordinate, the set of more general, inclusive words among those given by our sources, discarding those which are appropriate only for one item in the synset and not for the others. This will involve manual work.

On the other hand, proof of the possibility, in other cases, of an automatic synset construction was acquired for the word affetto (affection) belonging to the FEELING taxonomy.

A completely automatic search looking at the dictionary indications about synonymy, starting from affetto (affection)gave us the following:

affetto 0_1 sentimento di viva benevolenza verso una persona o una cosa SIN. affezione, amore, tenerezza affezione 0_1b tendenza ad amare SIN. affetto, attaccamento
amore 0_1c devozione (devotion)
attaccamento 0_0 vincolo affettivo SIN. affezione
tenerezza 0_2 (fig) sentimento di affetto delicato, talora misto a rimpianto SIN affetto

devozione 0_2a dedizione SIN affezione, fedeltá.
 devozione 0_2b attaccamento SIN affezione, fedeltá
 dedizione 0_0 il dedicarsi senza riserve a una persona , a un'attività etc.
 fedeltá 0_1 l'essere fedele SIN devozione

At this point by gathering the words explicitly indicated as synonyms affetto, affezione, amore, tenerezza, attaccamento, and the only synonym by definition type devozione, we obtain the synset:

{affetto0_1 (affection), affezione0_1b (fondness), amore0_1c (love), attaccamento 0_0 (attachment), tenerezza0_2 (tenderness), devozione 0_2a (devotion) dedizione 0_0 (devotion) fedeltá0_1 (faithfulness)}

This could probably be enlarged or restricted depending on the linguistic competence of each speaker, but is, in any case, an acceptable result. In fact, if we look at the WordNet 1.5 data, we find these synonyms for affection:

{affection, affectionateness, fondness, tenderness, heart, warmheartedness }

while love is indicated as hyperonym. Our automatically built synset does not map completely to this, but it seems to be a good initial approximation. Total identity may be difficult to reach due to differences in the different languages. Again in this case the problem is the hyperonym relation because we see that there are different hyperonyms or genus terms for this synset: sentimento (feeling), tendenza (bent), vincolo (tie), l'essere (to be), il dedicarsi (to devote oneself) and these do not fit well together in a single synset.

Furthermore it seems that the word devozione should probably be discarded from the synset because at this point a shift of meaning is noticeable and this item together with the items which follow in the chain, dedizione and fedeltá, should constitute another synset. This, obviously, will be task of the following (and manual) work of control and analysis.

II.2.5 Other semantic relations

We give here below some information on the distribution of the semantic relations within the noun subsets selected for the experiment.

FEELINGS

Our abstract nouns representing feelings are characterized by the high frequency of synonymy and antonymy, while the role relation does not appear at all and only one case of meronymy (set of) was found:

cenestesi: il complesso delle sensazioni....
 (the set of sensations..)

The Verb to Noun and Adjective to Noun relations are also frequent as this subset includes many derivatives, here below some examples are given.

Deverbal Nouns or Verb to Noun relations:

agitarsi agitazione (to agitate - agitation)

adorare adorazione (to adore - adoration)
 presentire presentimento (to have a presentiment of -presentiment)
 esultare esultanza (to exult - exultation)
 smarrirsi smarrimento (to lose one's way - confusion)

Deadjectival Nouns or Adjective to Noun relations:

tenero tenerezza (tender - tenderness)
 pesante pesantezza (heavy - heaviness)
 sereno serenità (serene - serenity)
 infelice infelicità (unhappy - unhappiness)
 allegro allegria (gay - gaiety)

INSTRUMENTS

The instruments subset is characterized either by the instrument role relation or by the meronymy. In fact most of the nouns in this subset are defined as "part of..", "set of.." or "instrument / tool / device .. used / employed / suitable for. ..". We found a very large number of instrument involvement relations such as:

attizzatoio : strumento con cui si attizza il fuoco
 (tool used to poke up the fire)
 trincetto : strumento del calzolaio...usato per tagliare il cuoio
 (tool used by shoemaker ..to cut leather)
 spettroscopio : apparecchio atto a studiare gli spettri luminosi
 (equipment to study luminous spectra)

We also found many meronymy relations of the type "set of" and "part of", e.g.:

cilindro: parte della macchina (part of a car)
 macchinario : complesso di macchine, meccanismi, apparecchi di una
 fabbrica (set of machines, mechanisms, equipment of
 a factory)

FOOD

The food subset gave just one relation of the involvement type:

salsa : condimento usato per arricchire le vivande..
 (dressing used to enhance the taste of food)

The most frequent relation in this subset is meronymy of different types, e.g.: part of, set of, but also quantity of, mixture of, element of, component of, made of etc..

A few examples from the subset are:

bistecca : fetta di carne di vitello cotta alla griglia
 rel: (portion of..)
 parmigiana : pietanza a base di melanzane, pomodoro e parmigiano
 rel: (made of..)
 crema : vivanda dolce a base di latte zucchero farina e uova

rel: (made of..)

grigliata : quantit=85 di cibo alla griglia

rel: (quantity of..)

insalata : mescolanza di cibi

rel: (mixture of..)

polpettone : impasto di carne trita e condimenti diversi

rel: (mixture of ..)

piatto : quantit=85 di cibo in esso contenuta

rel: (quantity of..)

pasto : l'insieme di cibi di cui ci si nutre quotidianamente

rel: (set of..)

mollica : la parte interna molle del pane

rel: (part of ...)

II.3 Detailed description of Spanish experiment

Experiments have been performed on FOOD&DRINKS, INSTRUMENT-1st level, and FEELING_1. Furthermore, as we had already extracted it, we include here the data for INSTRUMENT (complete taxonomy). Taxonomies have been extracted by automatic methods, hence they are subject to a ratio of mistakes.

We present here an estimation of the total number of relations potentially extractable from our taxonomy, and then the results for each individual experiments on that four subtaxonomies.

II.3.1 Estimation of the taxonomy as source of relations

Besides the experiments on specific subtaxonomies, an estimation of our general taxonomy as a potential source of semantic relations has been carried out. Our aim was to check whether the results from that concrete subtaxonomies could be generalized or not. These are the results:

	# Relations	# Noun senses	# Verb senses
Synonymy	8579	18736	1400
Antonymy	229	260	20
Meronymy:			
parte (part of)	2394		
conjunto (set of)	1680		
trozo (portion)	137		
pedazo (portion)	167		
masa (mass)	269		
mezcla (mixture)	215		
combinado (combination)	24		
miembro (member)	91		

For meronymy, the table shows meronymic relations appearing explicitly by means of specific patterns in dictionary definitions; it serves as a lower limit on the number of relations potentially extractable.

II.3.2 Results of the experiments

For each subset, we list the following data:

(i) productive nodes: these are the nodes of our taxonomies which have proved to be productive 'sub-tops' for the chosen subset.

#Spanish_node -- {wn1.5 corresponding synset} -- (wn1.5 gloss)

(ii) number of nodes (senses): total number of spanish taxonomy nodes, i.e. word senses, extracted as belonging to the taxonomy of the corresponding subset. This number is also shown distributed by productive sub-tops.

(iii) taxonomy depth: maximal and most frequent length - number of nodes - of the taxonomy extracted.

(iv) taxonomy membership precision: Percentage of extracted synsets correctly assigned to the taxonomy. This data is obtained looking at a random sample of 100 elements. It means that, for instance, 100% of the nodes of the taxonomy extracted as belonging to INSTRUMENT-1st are actually instruments (Note that the percentage, in this case, is very high: this is due to the fact that this first-level subset has been extracted simply by retrieving all word-senses in the MRD whose definition genus is instrumento_1_1).

(v) coverage with bilingual: data related to automatic linking to WN1.5 (accuracy of our automatic procedures here is still under evaluation).

a) <subset> synsets: number of synsets in WN1.5 for the corresponding subset

b) <subset> synsets using bilingual dictionary: percentage of synsets in a) that have been linked to some spanish word sense using bilinguals.

c) <subset> synsets connected to <subset> tax senses: number of nodes-senses (and number of words for that nodes) in the spanish taxonomy automatically linked to WN1.5, and number of target nodes in WN1.5 for that links that also belong to the subset.

(vi) other relations: number of links other than hyponymy-hyperonymy which have been extracted from the spanish MRD for the taxonomy obtained.

Experiment 1: FOOD & DRINKS

(i) productive nodes: 12

#caldo -- {broth,stock} -- (liquid in which meat and vegetables are simmered; used as a basis for e.g. soups or sauces)

#alimento -- {food,nutrient} -- (any substance that can be metabolized by an organism to give energy and build tissue)

#bebida -- {beverage,drink,potable} -- (any liquid suitable for drinking)

#calabaza -- {pumpkin} -- (usu large púlpny deep-yellow round fruit of the squash family maturing in late summer or early autumn)

#tocino -- {porc,pork} -- (meat from a pig)

#carne -- {meat} -- (the flesh of animals used as food)

#comestible -- {grocery,foodstuff} -- (commodities sold by a grocer)

#grano -- {grain,food grain,cereal,cereal grain} -- (farinaceous grain suitable for food)

#planta -- {plant,flora,plant life} -- (a living organism lacking the power of locomotion)

#pan -- {bread,breadstuff,staff of life} -- ([no gloss in WN1.5] edible baked good made of flour etc.)

#harina -- {flour} -- (finely ground meal of any cereal grain usu. freed from bran)

#vianda -- {food,nutrient} -- (any substance that can be metabolized by an organism to geive energy and build tissue)

(ii) **number of nodes** (senses): 3339

distributed by subtaxonomies:

Subtaxonomy	# Nodes
caldo	18
alimento	418
bebida	1498
calabaza	33
tocino	42
carne	240
comestible	430
grano	166
planta	271
pan	185
harina	35
vianda	103

(iii) **taxonomy depth:**

maximal: 8

most frequent: 3

(iv) **taxonomy membership precision:** 81%

(v) **coverage with bilingual:**

food synsets: 2290

food synsets using the bilingual dictionary: 618

416 Spanish food senses (261 words) connected to

286 wn1.5 food synsets

(vi) **other relations:**

synonymy (senses with at least 1 synonym): 105
antonymy: 0
involvement: 68
part-of: 40
set-of: 10
has_substance: 33

Experiment 2: INSTRUMENT (First Level)

(i) **productive nodes:** 1

#instrumento -- {instrument} -- (a device that requires skill for proper use)

(ii) **number of nodes** (senses): 568

(iii) **taxonomy depth:** 1

(iv) **taxonomy membership precision:** 100%

(v) **coverage with bilingual:** No data available

(vi) **other relations:**

synonymy: 28
antonymy: 0
involvement: 97
part-of: 1
set-of: 14
has substance: 0

Experiment 3: INSTRUMENT (All levels)

(i) **productive nodes:** 4

#instrumento -- {instrument} -- (a device that requires skill for proper use)
#herramienta -- {tool} -- (an implement used in a practice of a vocation)
#varilla -- {rod} -- (a long thin implement made of metal or wood)
#mecanismo -- {device} -- (an instrumentality invented for a particular purpose)

(ii) **number of nodes:** 3263

instrumento 2544
herramienta 128
varilla 82
mecanismo 509

(iii) taxonomy depth

maximal: 10

most frequent: 2

(iv) taxonomy membership precision: 87%**(v) coverage with bilingual:**

artifact synsets: 9109 (no data extracted for 'instrument' in
WN1.5; just for the more general 'artifact')

artifact synsets using the bilingual dictionary: 3714

872 Spanish 'instrumento' senses (623 words) connected to

1398 WN1.5 artifact synsets

Experiment 4: FEELING_1**(i) Productive nodes: 11**

#cólera -- {anger,choler,ire} -- ([no gloss in WN1.5] strong feeling against something or someone)

#placer -- {pleasure,pleasance} -- (a fundamental feeling that is hard to define but that people desire to experience)

#confianza -- {confidence} -- (a feeling of trust in someone or something)

#contento -- {happiness} -- (emotions experienced when in a state of well-being)

#pesar -- {sorrow} -- (the emotion of great sadness)

#deseo -- {desire} -- (the feeling that accompanies an unsatisfied state)

#empeño -- {resoluteness,firmness,resolve,resolution} -- (firmness of purpose)

#júbilo -- {joy,joyousness,joyfulness} -- (the emotion of great happiness)

#pasión -- {passion,passionateness} -- (strong feeling or emotion)

#perturbación -- {disturbance,perturbation,upset} -- (an unhappy and worried mental state)

#sentimiento -- {feeling} -- (the psychological feature of experiencing affective and emotional states)

(ii) **number of nodes** (senses): 632 distributed in subtaxonomies as follows:

Subtaxonomy	# Nodes
cólera	11
placer	27
confianza	20
contento	12
pesar	6
deseo	67
empeño	82
júbilo	7
pasión	136
perturbación	195
sentimiento	69

(iii) **taxonomy depth:**

maximal: 10

most frequent: 2

(iv) **taxonomy membership precision:** 75%

(v) **coverage with bilingual:**

feeling synsets: 397

feeling synsets using the bilingual dictionary: 267

136 Spanish sentimiento senses (117 words) connected to 116 WN1.5 feeling synsets

(vi) **other relations:**

synonymy: 33

antonymy: 1

involvement: 4

part-of: 0

set-of: 1

has_substance: 0

Appendix III Experiments on discovering interlingual synonymy.

One of the main steps towards building the EWN database is the linkage of every monolingual wordnets to WN1.5. The latter acts as an interlingual index that connects every pair of languages in the database.

FUE, on the other hand, is going to build the spanish wordnet by using bilingual MRDs, and other sources, to connect spanish word forms to wn1.5 synsets. FUE experiments carried out to test this approach may well serve to foresee the problems that all partners will face when linking their monolingual wordnets to WN1.5.

The primary information explicitly available from bilingual MRDs are word to word pairs (english to spanish and viceversa). The words in that pairs, however, are not disambiguated. Therefore, when one or both words in a pair are polysemic, the pair could be insufficient to establish an accurate link. In such cases, we need contextual information to disambiguate the meanings of the words involved; that kind of information may come from additional translation words, semantic cues, definitions, etc, combined with taxonomies and other information extracted from monolingual MRDs.

FUE strategy is to start extracting all possible links without considering contextual information, and in a second step to use context to disambiguate the remaining pairs. The criteria to establish direct links and the methodology to exploit contextual information are based on (Rigau and Aguirre, 95).

This appendix shows the results of a preliminary experiment to extract synonymy links using only word to word mappings of the VOX bilingual dictionary (~ 15.000 entries per language), without considering any contextual information. According to the degree of ambiguity of both words in a pair, eight different criterion are considered from which a link can be extracted.

In the experiment, all pairings fitting any of the eight criteria are automatically treated to extract the links. For each criterion, a random subset of 10% of the links is manually verified to test its accuracy. A software environment has been created to apply automatically such criteria and to facilitate manual verifications.

These are the eight criteria:

Criterion 1

Conditions:

- Spanish word (SW) has only 1 translation in bilingual MRD
- English word (EW) has only 1 translation (SW) in bilingual MRD
- EW is monosemous in WN1.5 (belongs only to 1 synset)

Link: The spanish word is linked to the only WN1.5 candidate synset.

Criterion 2

Conditions:

- SW has more than 1 translation ($EW_1, EW_2 \dots EW_n$)
- EW_i has only 1 translation (SW)
- EW_i is monosemous in WN1.5

Link: The spanish word SW is linked to the only WN1.5 candidate synset (one link for each translation EW_i that satisfies the conditions).

Criterion 3

Conditions:

- SW_1, SW_2, \dots, SW_n 1 translation (EW)
- $EW > 1$ translation (SW_i for some i)
- EW monosemous

Link: All spanish translations SW_i of the english word are attached to the only WN1.5 candidate synset.

Criterion 4

Conditions:

- SW 1 translation (EW)
- EW 1 translation (SW)
- EW polysemous

Link: The spanish word SW is linked to every WN1.5 synset in which the english word appears.

Criterion 5

Conditions:

- SW 1 translation (EW)
- $EW > 1$ translation (SW, $SW_1, SW_2, \dots SW_n$)
- EW polysemous

Link: The spanish word SW is linked to every WN1.5 in which the english word appears. This criterion is included for test purposes, but obviously leads to erroneous assignments in some cases. For example, the spanish word "primavera", which means "spring" as a season, is attached also to "spring" as elastic metal device, as fountain, etc.

Criterion 6

Conditions:

- SW > 1 translation ($EW_1 \dots EW_n$)
- EW_i 1 translation (SW) for some i
- EW_i polysemous

Link: The spanish word SW is attached to all WN1.5 synsets that contain EW_i , for every EW_i that satisfies the conditions above.

Criterion 7

Conditions:

- Same as criterion 5, plus
- two EW belonging to the same WN1.5 synset have the same SW as translation

Link: such spanish word is mapped to that synset.

Criterion 8

Conditions:

- Same as criterion 6, plus
- Besides translation, the EW has a field identifier whose value is an english word EW_2 that belongs to the same WN1.5 synset than EW.

Link: Spanish word SW is attached to the WN1.5 synset that contains EW and EW_2 .

A table with the results of the manual verification for each criterion follows.

For each criterion, the test has been performed on 10% of the set of the total mappings obtained using such criterion. No results are shown for criterion 5, as the number of errors was too high and it was rejected before finishing the verification.

Criterion	# LINKS	% OK	% KO	% HYPO	% HYPER	% NEAR
1	2680	93.2	3.3	0.7	0.7	1.8
2	2520	88.5	2.0	0.4	4.7	4.4
3	1470	87.7	7.5	2.0	1.3	1.3
4	2930	79.1	13.0	0.3	0.7	6.8
5	***	***	***	***	***	***
6	3810	72.7	19.4	0.0	0.5	7.3
7	3300	85.8	5.1	0.6	1.8	6.7
8	820	78.0	8.5	2.4	2.4	8.5

#LINKS: Number of links established. Only sets 7 and 8 have some overlapping with 6; the rest are disjunct.

OK: correct links

KO: fully incorrect links

HYPO: links to a hyponym of the correct synset.

HYPER: links to a hyperonym of the correct synset.

NEAR: links to near synonyms that could be considered OK.

Most of the criteria offer quite reliable results. In fact only the fifth criterion is absolutely rejectable. In the other cases, either taken isolated or combined with additional restrictions, the criteria seem to be accurate enough.

The overall number of links obtained (compared to the number of entries in the dictionary, about 30,000) leads us to estimate that about an 80% of the noun synsets for spanish could be extracted that way.

REFERENCES

- Cruse, D.A. *Lexical Semantics*. Cambridge University Press. Cambridge, 1986.
- Martin W. G.A.J. Tops. *Groot woordenboek Engels-Nederlands*. Utrecht: Van Dale Lexicografie. 1984.
- Miller, G.A. and C. Fellbaum. Semantic Networks of English. In Levin B. and Pinker S. (eds.) *Lexical and Conceptual Semantics*. Elsevier Science Publishers. Amsterdam, 1992.
- Miller, G, R. Beckwith, C. Fellbaum, D. Gross and K. Miller *Five papers on WordNet*. CSL Report 43. Cognitive Science Laboratory. Princeton University. 1990.
- Rigau, G. and Aguirre, E.: Linking bilingual dictionaries to Wordnet. In *Proceedings of the Workshop on The Computational Lexicon*, Barcelona 1995.
- Sterkenburg P. van - W.J.J. Pijnenburg *Groot woordenboek van hedendaags Nederlands*. Utrecht: Van Dale Lexicografie. 1984.
- Winston M, R., Chaffin and D. Herrmann. A taxonomy of Part-Whole relations. *Cognitive Science* 11, pp. 417-444. 1987.