

# Coding the Semantic Relations For Basic Nouns and Verbs

Piek Vossen , Laura Bloksma, *Computer Centrum Letteren, University of Amsterdam*  
Antonietta Alonge, Francesca Bertagna, Adriana Roventini, Elisabetta Marinai, *Istituto di  
Linguistica Computazionale, CNR, Pisa*

Carol Peters, *Istituto di Elaborazione della Informazione, CNR, Pisa*

Wim Peters, *University of Sheffield*

Salvador Climent, *University of Barcelona*

Version 4, April 19, 1997

Final



**Deliverable D010, D011, WP3.2, WP4.2,  
EuroWordNet, LE2-4003**

This document is a joint product of the University of Amsterdam, The Istituto Linguistica di Computazionale del CNR at Pisa, The University of Barcelona, The Politecnica de Catalunya Barcelona, UNED Madrid and the University of Sheffield.

Identification number	LE-4003-D-010 LE-4003-D-011
Type	Document
Title	Coding the Semantic Relations For Basic Nouns and Verbs: building the first subset.
Status	Final
Deliverable	D010, D011
Work Package	WP3 & WP4
Task	T3.2 & T4.2
Period covered	October 1996- January 1997
Date	April 19, 1997
Version	4
Number of pages	127
Authors	<ol style="list-style-type: none"> <li>1. Piek Vossen , Laura Bloksma, <i>Computer Centrum Letteren, University of Amsterdam</i></li> <li>2. Antonietta Alonge, Francesca Bertagna, Adriana Roventini, Elisabetta Marinai, <i>Istituto di Linguistica Computazionale, CNR, Pisa</i></li> <li>3. Carol Peters, <i>Istituto di Elaborazione della Informazione, CNR, Pisa</i></li> <li>4. Wim Peters, <i>University of Sheffield</i></li> <li>5. Salvador Climent, <i>University of Barcelona</i></li> </ol>
WP/Task responsible	FUE/PSA/AMS/SHE

Project contact point	Piek Vossen Computer Centrum Letteren University of Amsterdam Spuistraat 134 1012 VB Amsterdam The Netherlands tel. +31 20 525 4669 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	Multilingual Lexical Semantic database, Information Retrieval, Language Engineering.
Abstract	This document reports on the first building phase in which the semantic relations are encoded for EuroWordNet. The work involved the definition of a set of common Base Concepts: the meanings that are most important in the wordnets. Furthermore for each site the progress on building of the relations for these concepts is described. This report will be updated when the work for the first subset is finished (September 1997).
Status of the abstract	Final
Received on	
Recipient's catalogue number	

## Executive Summary

This document reports on the first building phase in which the semantic relations are encoded for EuroWordNet. The work involved the definition of a set of common Base Concepts and for each site the progress on building of the lexical semantic relations for these concepts. The selection of the Base Concepts proceeded as follows:

- separate selections of those meanings having most relations (as a hyperonym or as a genus word) and the most important position (as a top or close to the top) in the hierarchies in the available resources in each of the languages
- translation of these Base Concepts to the most close WordNet1.5 synset
- the intersections of these selections

Because these Base Concepts have most relations and/or occupy the major positions in the hierarchy they can be seen as the core of the wordnets. The resulting set of 628 WordNet1.5 synsets is thus included in the starting set by all the partners. Furthermore, the manual work is concentrated on these meanings guaranteeing high quality and compatibility for the same core of all the wordnets. The semi-automatic techniques will then be used for the extensions to more specific areas of the vocabulary. To further coordinate the work on these Base Concepts we have developed a first version of a top-ontology to cluster the concepts into different types. These clusters are used to measure progress and discuss related problems. The top-ontology is based on other ontologies (WordNet1.5, Acquilex, Sift) and extended to represent the variety of meanings represented by the Base Concepts. Finally, this document contains descriptions and tables on the coding of the relations for these cores with some first extensions for each of the building sites.

The results cover the work on the First Subset up to February 1997 and will continue until September 1997 after which it will be verified. The final results for the First Subset will be described in the deliverables D014, D015 and D017, which can be seen as updates of this document.

# Table of Contents

1. Introduction.....	8
2. Agreed building strategy .....	9
3. Base Concept Selection and Top Ontology clustering.....	12
3.1 The set of common Base Concepts.....	13
3.2. Top-Ontology clustering of the Base Concepts.....	15
4. First Building Phase AMS.....	21
4.1 Tools and resources .....	21
4.2 AMS approach.....	23
4.3 AMS BC selection .....	25
4.4 AMS Results for the first subset .....	28
4.5 Spent resources and prospects for the extendibility .....	38
5. First Building Phase PSA .....	40
5.1 Tools and Resources .....	40
5.1.1 MRDs and Lexical Databases as input .....	40
5.1.2 Local Tools and Technology .....	41
5.2 Local approach .....	43
5.3 PSA BCs selection .....	43
5.3.1 Criteria and methods adopted for the preliminary selection .....	43
5.3.2 Treating ‘Missing’ concepts and ‘Drop-outs’ .....	44
5.4 PSA results for the first subset .....	47
5.4.1 Results obtained for verbs .....	48
5.4.2 Results obtained for nouns.....	50
5.5 Looking for heuristics for automatic encoding.....	54
5.6 Prospects for the extendibility .....	56
6. First Building Phase. FUE.....	57
6.1 Tools and Resources .....	57
6.1.1 Databases as Input.....	57
6.1.2 Local Tools and Technology .....	58
6.2 Local Approach .....	59
6.2.1 Nouns .....	60
6.2.2 Verbs .....	62
6.2.3 Work on extending and refining the data.....	63
6.2.4 Top-Ontology Clustering.....	63
6.3 FUE BC Selection.....	64
6.4 FUE Results.....	64
6.4.1 Figures .....	64
6.5 Amount of resources spent.....	68
6.6 Quality of the data. Evaluation. ....	68
6.7 Prospects for the extendibility .....	71
7. First Building Phase Sheffield.....	72
7.1 Tools and Resources .....	72

7.2 Local Approach .....	73
7.3 SHE Base Concepts selection.....	74
7.4 Results.....	76
7.5 Prospects for the extendibility. ....	77
8. Conclusions .....	78
References .....	78
EuroWordNet Deliverables referred to.....	79
Appendix I Nominal Base Concepts selected by all four sites.....	80
Appendix II Verbal Base Concepts selected by all four sites .....	83
Appendix III Top-Ontology concepts.....	85
Appendix IV Hierarchy of EuroWordNet Top-Concepts, version 2.....	92
Appendix V Hierarchy of Top-Concepts with the clustering of Base Concepts, version 294	
Appendix VI Genuine Drop-Outs AMS .....	115
Appendix VII FUE-Sample of the Noun database (hyponymy).....	118
Appendix VIII FUE-Sample of the Verb database (hyponymy).....	120
Appendix IX FUE: working out the ‘communication’ semantic field .....	122

## List of Tables

Table 1: Original Selection of Base Concepts represented by WordNet1.5 equivalences	13
Table 2: Base-Concept Intersection-Pairs .....	14
Table 3: Base-Concepts, combined intersections .....	14
Table 4: Missing, selected and non-selected Base-Concepts per site .....	15
Table 5: Number of entries and senses in VLIS database (AMS) .....	22
Table 6: Conversion of VLIS relations (AMS) .....	23
Table 7: Dutch Initial Base Concepts, measured in WordNet1.5 senses (not synsets).....	25
Table 8: Dutch Initial Base Concepts, measured in VlisSenses:.....	26
Table 9: Dutch Drop outs, measured in VlisSenses.....	26
Table 10: Dutch Drop-Outs as Hyponyms of Selected BCs, measured in VlisSenses .....	26
Table 11: Final Set of Base Concepts measured in Vlis senses (AMS) .....	27
Table 12: First Subset differentiated for Core and Core + 1 <sup>st</sup> . Extension (AMS) .....	29
Table 13: Relations Encoded for First Subset (AMS) .....	32
Table 14: Copied and added relations for 798 BC-Nouns (AMS) .....	34
Table 15: Copied and added relations for 251 BC-Verbs (AMS). .....	35
Table 16: Copied and added relations for total First Subset Nouns and Verbs (AMS).....	35
Table 17: Copied and added relations for other Parts of Speech (AMS). .....	36
Table 18: Major Semantic Fields of AMS First Subset in terms of hyponyms.....	36
Table 19: Effort Spent for building the AMS First Subset with extensions .....	38
Table 20: Remaining Effort Spent for Extension of the concept “gebouw” (building) .....	38
Table 21: Estimated Effort for the remaining work (AMS). .....	39
Table 22: Entries, Senses and Examples in the Italian LDB .....	40
Table 23: Taxonomic Data represented in the Italian LDB .....	41
Table 24: Italian Initial Base Concept Synsets (PSA).....	44
Table 25: Relations Encoded for the First Subset Verbs (PSA).....	49
Table 26: Relations Encoded for the First Subset Nouns (PSA).....	54
Table 27: Defining Formulae for Agents/Patients (PSA).....	56
Table 28: Number of entries and senses (FUE).....	58
Table 29: Taxonomic data per sense (FUE).....	58
Table 30: Noun First Subset differentiated for Core and Core + 1 <sup>st</sup> . Extension (FUE).....	64
Table 31: Verb First Subset differentiated for Core and Core + 1 <sup>st</sup> . Extension (FUE).....	64
Table 32: First Subset differentiated for Core and Core + 1 <sup>st</sup> . Extension (FUE) .....	64
Table 32: Hierarchical overview of distribution of sense per (main) cluster (FUE) .....	66
Table 33: Final Set of Base Concepts (SHE) .....	76

## 1. Introduction

This report describes the building of the first wordnet subsets for nouns and verbs in the EuroWordNet project (LE2-4003). It represents the deliverables D010 (Build the first noun subset) and D011 (Build the first verb subset). The actual work described in this document covers a period of three months (from the 1<sup>st</sup> of November 1996 until the 1<sup>st</sup> of February 1997). The results are therefore very preliminary, only illustrating the implementation of the design and specification phase of the project.

As described in the Technical Annex (TA) of EuroWordNet (D000), the building of the resources takes place in two phases:

- 1) first we will produce the correct language-internal structures, mainly from a monolingual perspective,
- 2) and in the second phase compare and, if necessary, restructure the monolingual wordnets or the equivalence relations.

The first task is mainly achieved using local tools and resources and different approaches. The latter will be done using the Novell tools and the EuroWordNet database, which is specifically designed for comparing the resources. Currently, the work has been restricted to the first phase only. No comparison or consequently restructuring has been done yet.

The actual building of the resources follows an intense period of discussion and study in which the multilingual database has been designed, the covered vocabulary has been specified and the relations have been defined. The specification of the database is broader than the scope of actual data that are being produced (see D001, D005, D006 and D007 for details and motivation of the design). Given the fact that the available resources and budget of the project are limited, it is not feasible to comprehensively code all the possible relations and relation properties. Furthermore, also the resources and tools being used by the different sites are different. Where AMS makes use of a database of Van Dale in which some of the relations are already explicitly coded between senses, PSA uses several resources which are combined and which still have to be disambiguated manually to achieve these relations. FUE again follows a completely different strategy where WordNet1.5 is first translated into Spanish and then post-processed and SHE directly makes use of WordNet1.5 for English, combining it with other resources such as LDOCE (Procter 1978) and the Celex database. This gives each site a different starting point, the details of which are further described below. It will be clear that these differences also lead to differences in coverage and productivity. Not all aspect can be provided by all partners but the merge of all that can be provided still requires a broad design.

The structure of this report is then as follows. In section 2 we shortly describe the general building strategy that we have adopted. The building experiments carried out in the specification phase showed that it was necessary to take some specific measures to ensure compatibility of the wordnets built at the different sites with different resources and tools.



Section 3 then contains a specification of the first subset that is being built. The subset is described using an ontology of language-independent top-concepts. To build this ontology some of the work of WP5 has been carried out earlier. The selection of Base Concepts has been clustered around the classes of this top-ontology. Sections 4, 5, 6 and 7 then describe the building progress at each site. Section 8 contains some preliminary conclusions.

## 2. Agreed building strategy

As stated above, the aims of the project can be divided into two general tasks:

- the creation of monolingual wordnets for the most general parts of the vocabulary in several European languages.
- to combine these wordnets in a multilingual database.

The latter aim, which includes the first and is more ambitious, has some very interesting possibilities:

- to compare the independently-created resources so that inconsistencies can be traced and language specific properties are elicited.
- to share language-independent information.
- to more easily update and customize the resources via the language-independent information.
- to use it for multilingual information retrieval.

These benefits are further explained and motivated in D001 and D007. Benefits often have a price, in this case it means that the individually created wordnets can only be combined in a useful and proper way when the information coded in each is compatible. There are two kinds of compatibility issues (see Vossen 1996):

- to ensure that we cover the same vocabulary and still work relatively independently.
- to ensure that we apply the same interpretations when coding the relations.

To anticipate such problems we have carried out two experiments (the experiments are described in more detail in D005 and D006). The first experiment was set up to determine the vocabulary that will be covered in the wordnets. The specification of the vocabulary was faced with two conflicting requirements:

- 1) there must be sufficient overlap in the vocabularies across the different sites
- 2) the vocabularies should leave room for including language-specific lexicalizations and structures

The first requirement could easily be achieved by taking a selection of WordNet1.5 synsets that are translated into each of the languages. However, this would mean that any language specific lexicalization and structure that deviates is blended by imposing the English selection. We therefore decided to determine the selection more globally on conceptual grounds.<sup>1</sup>

To see to what extent it is possible to globally classify the vocabulary each site has defined the major concepts or so-called Base Concepts (henceforth BCs) in their resources. The selection of a common set of BCs will provide us with a common starting point for extending the wordnets. In that way we can assure that the same main fields are covered at all sites without depending too much on each other in progress and without fully predetermining the selection in advance. Each side has then selected those word meanings that have most relations and occupy the highest positions in the hierarchies (as far as they have been structured) as their choice of BCs. These selections have been translated into equivalent WordNet1.5 synsets. These WN1.5 synsets have been exchanged and compared. From this it followed that the intersection (expressed in WN1.5 equivalences) in selected BCs was extremely low (only 4 up to 10%). This clearly indicated that we have to be very cautious determining the coverage.

The second experiment involved the coding of the relevant relations for some areas of the vocabulary. Ultimately, the findings provided the basis for specifying the EuroWordNet relations as described in D005 and D006. However, the experiment also showed some apparent problems. First of all it appeared that creating the fundamental relations: synonymy and hyponymy, was very complicated. Judging synonymy or hyponymy was subjective and time-consuming while the available resources are not very helpful or reliable. Furthermore, the definitions and test for the EuroWordNet relations seem to work for the more coarse-distinctions but not for closely-related words at more abstract levels which have many vaguely-distinguished meanings. Secondly, it became clear that establishing equivalence relations with WordNet1.5 was more complicated than expected. The main reasons for this are:

---

<sup>1</sup> This approach of starting from WordNet1.5 is nevertheless pursued by one of the sites: FUE. However, FUE will verify their approach with an independent Spanish resource, which will give us useful information on the results of different approaches.

- the extreme sense-differentiation in WN1.5: a verb such as “clean” has 19 senses differentiated with respect to the kind of substance or objects that are removed or cleaned.
- inconsistencies in coding senses across the resources. For example, regular polysemy is coded in different ways within and across resources: by coding all senses expressed, by coding any subset of all possible senses, and by collapsing senses in a single sense.
- in WordNet1.5 many extra levels are introduced in the hierarchy. These involve both technical classifications and artificial non-linguistic levels used to group particular hyponyms. Such levels are not expected in a normal dictionary.

A positive conclusion of the experiment was that these problems are restricted to the higher, more abstract levels of the hierarchy where definitions appear to be vague and words tend to be polysemous. The construction of more specific areas can be done on a semi-automatic basis, causing less problems. We therefore decided to divide the building work in two separate tasks:

- 1) the coding of what we call the Base Concepts of each resource
- 2) the coding of so-called vertical extensions of these Base Concepts

The coding of the Base Concepts will mostly be done manually where we will try to get a maximum of overlap and consensus in approach. These meanings occupy major hierarchical positions and/or have a large number of relations. They therefore play a major role in the database and it is very crucial that coverage and interpretation is the same across the sites. To achieve consensus with respect to the building of the wordnets for the Base Concepts we have taken the following measures:

- we have created explicit tests for all the relations in the different languages with examples for each of these relations and the use of labels for these relations (D005 and D006).
- we are planning to exchange the definition patterns and heuristics for (semi-)automatically deriving the data:
  - *we use similar heuristics for the extension,*
  - *provide a kind of general semantic context defining the relation.*
- each group will produce a manual or guide for producing wordnets with these relations. In these guides we will give an overview of all the problems encountered and what the agreed solutions are taken by all the sites.
- after the data have been loaded into the common EuroWordNet database, comparing the results will give many clues and suggestions for restructuring the first results later on. Such comparison will show (Vossen 1996, Vossen et al 1997):
  - *conflicting relations across wordnets*
  - *alternative classifications*
  - *mismatches in equivalence relations*

The extension can be done semi-automatically because meanings are more clearly defined and less polysemous.

The work reported in this document mainly describes the work done for these Base Concepts. The work is still in progress and will be continued. In addition to the Base Concepts we have vertically extended some of areas to measure the efforts for enlarging the coverage at more specific levels. By comparing the results and efforts we thus get a first idea of the total costs involved. The following extensions have initially been selected:

1<sup>st</sup> order concepts:

- human occupations
- places
- instruments
- food & drinks
- buildings

2<sup>nd</sup> order entities

- physical states or properties (shape, weight, colour, size, taste, etc.)
- communication
- perception
- feelings
- movements
- know/knowledge
- change in physical state

The actual coverage of these extensions may differ from site to site. What is actually coded also depends on the approach followed and the tools and resources that are available. Close co-operation is less crucial for the extensions since consensus and compatibility is mostly determined at the Base Concept level. Obviously, this work is still in progress too.

In the next section we will first describe how the final set of Base Concepts has been determined and how the work has been divided into coherent clusters. Sections 4, 5, 6 and 7 describe the individual results of building the first subset at each site, where the different starting points, resources and tools are outlined.

### **3. Base Concept Selection and Top Ontology clustering**

The first step in creating coherent wordnets across the sites involved the selection of a set of Base Concepts that represent the most important concepts in each resource and will be used by all sites as the common starting point. The specific selections by each site have been described in D005 and D006. These selections have been compared and merged to get the set of common Base Concepts.

Next these Base Concept have been clustered by creating an artificial language-independent top-ontology. This clustering functions as a first basic classification of concepts, so that we can exchange progress, ideas and problems per cluster and can start discussing some basic semantic distinctions.

### ***3.1 The set of common Base Concepts***

All the sites have sent the list of WordNet synsets that occurred as translations of their Base Concepts to AMS. These lists contained the file offset position with a part of speech tag, e.g.: 00016649-n and 01472320-v, which are unique WordNet1.5 IDs for the synsets of the nouns “act 1” and the verb “be 4” respectively. To determine the intersection of these lists we have loaded these IDs into the AMS lexical database (see section 4 below). This database uses a more traditional sense-based entry structure rather than the synset structure of WordNet1.5 which means that the figures are expanded. In order to be able to compare the selections the synsets had to be converted to WordNet senses via an index:

*Table 1: Original Selection of Base Concepts represented by WordNet1.5 equivalences*

Sites	Synsets in WN1.5	AMS LDB Senses in WN1.5
<b><i>Nouns</i></b>		
AMS	1019	2265
FUE	362	673
PSA	315	700
SHE	1090	2000
Total Non-unique	2786	5638
<b><i>Verbs</i></b>		
AMS	321	1174
FUE	105	354
PSA	97	344
SHE	197	644
Total Non-unique	720	2516

For these senses we have determined the intersection for each language pair:

*Table 2: Base-Concept Intersection-Pairs*

	Total	AMS		FUE		PSA		SHE	
<b>Nouns</b>									
AMS	2265	X	X	210	9.2%	398	17.5%	596	26.3%
FUE	673	210	31.2%	X	X	96	14.2%	222	32.9%
PSA	700	398	56.8%	96	13.7%	X	X	321	45.8%
SHE	2000	596	29.8%	222	11.1%	321	16.0%	X	X
<b>Verbs</b>									
AMS	1174	X	X	142	12.0%	143	12.1%	250	21.2%
FUE	354	142	40.1%	X	X	65	18.3%	57	16.1%
PSA	344	143	41.5%	65	18.8%	X	X	111	32.2%
SHE	644	250	38.8%	57	8.8%	111	17.2%	X	X

If these intersections are combined we get the following results for nouns and verbs:

*Table 3: Base-Concepts, combined intersections*

<b>Union of all Noun senses</b>	<b>4247</b>	
Intersection of all	81	
Union of Intersection Pairs	999	497 synsets
Non-intersecting senses	3248	
<b>Union of all Verb senses</b>	<b>1873</b>	
Intersection of all	18	
Union of Intersection Pairs	471	131 synsets
Non-intersecting senses	1402	

We see that the intersection of all selections is low: 81 noun senses and 18 verb senses. These are listed in Appendix I and II respectively. We therefore decided to select all senses that occur in at least two selections: the union of the intersection pairs. The final set of common Base Concepts which is used by all sites as a starting point thus consists of 497 noun synsets (999 senses) and 131 verb synsets (471 senses) in WordNet1.5.

When compared to the original selection of Base Concepts of each site, each group has to deal with the following divisions:

Table 4: Missing, selected and non-selected Base-Concepts per site

Site	POS	Total of AMS WN senses	Drop- Outs	Missing	Selected
AMS	Nouns	2265	1441	175	824
	Verbs	1174	737	34	437
FUE	Nouns	673	369	695	304
	Verbs	354	178	295	176
PSA	Nouns	700	206	505	494
	Verbs	344	178	305	166
SHE	Nouns	2000	1232	231	768
	Verbs	644	344	171	300

Each site has then extended their selected senses with equivalences for their *missing* wordnet senses. Furthermore, each site has inspected the *drop-outs* to verify the selection criteria and to see whether crucial meanings have been left out. The latter are added and exchanged with the other sites. So for AMS this means that 175 nouns and 34 verbs synsets in WordNet are missing in the Dutch selection and have to be translated to Dutch senses. Furthermore, AMS has inspected the Dutch senses corresponding with the 1441 noun synsets and the 737 verb synsets (the AMS drop-outs) for re-selection as Base Concepts. In the sections on each site's progress the results of this are described.

Ultimately, each site has thus determined a selection of Base Concepts in each language which overlaps with the other sites in terms of 497 noun synsets and 131 verb synsets in WordNet1.5. This common core is used to achieve consensus and overlap in coverage across the sites.

### 3.2. Top-Ontology clustering of the Base Concepts

To get to grips with this set of Base Concepts we have constructed a top-ontology of 87 concepts to which each of the Base Concepts has been related. The ontology was initially based on the top-classifications in WordNet1.5, Aquilex (BRA 3030, 7315), Sift (LE-62030, Vossen and Bon 1996) and Aktions-Art models. Next the above Base Concepts (497 Noun synsets and the 131 Verb synsets) have been classified by AMS by linking them to the closest and most relevant ontology class. This has led to some restructuring of the ontology to represent the diversity of concepts. For new areas for which there was no proper classification new nodes have been created and positioned in relation to the other nodes.

Both the top-ontology and the Base Concept clustering has been imported in the Novell ConceptNet1.0 database (Cuyppers and Díez-Orzas 1996). This first version of the top-ontology and the Base Concept clustering has been distributed to the other sites. They have verified the ontology and checked the clustering with the equivalence of the Base Concepts in their local languages. This verification has again led to a restructuring of the ontology and the clustering. The second version of the top-ontology and the base concept

clustering was then released at the 2<sup>nd</sup> of December 1996. This version was used by all the sites to start the building of the first subset.

In Appendix III an alphabetic listing is given of these Top-Concepts with a definition and/or some examples and with a link to its superordinate concept. In some cases specific opposition relations between concepts have been specified by listing the disjoint concepts, e.g. “AnimateObject” is disjoint with “InanimateObject”. In Appendix IV a hierarchical overview is given of the top-ontology.

The top is represented by Top which is subdivided into two disjoint classes:

FirstOrderEntity	any concrete entity perceivable by the senses which is not a state, relation or an event; e.g. thing, anything
HighOrderEntity	any property, static relation or dynamic change (which cannot be grasped, heard, seen, felt as an independent physical thing); e.g. continue, occur, apply, take place, event, situation, state.

The class FirstOrderEntity is then subdivided into 9 more specific classes of which only Object and Substance form disjoint classes:

Functional	anything with a particular purpose but diverse denotation. Typically it can be used for nouns that can refer to any substance, object which is involved in a certain way in some event or process; e.g. remains, product, piece of art	can be combined with any constitutional class
Group	any concrete entity consisting of multiple discrete objects (either homogeneous or heterogenous sets), typically people, animals, vehicles; e.g. traffic, people, army, herd, fleet	
Object	any conceptually-countable concrete entity with an outer limit; e.g.	Disjoint=Substance
Part	anything which is contained in an object, substance or a group; head, juice, nose, limb, blood, finger, wheel, brick, door	can be combined with object and substance
Place	concrete entities functioning as the location for something else; e.g. place, spot, centre, North, South	
Portion	a limited amount of a substance which can be an independent object or a part; e.g. drink, splash, piece, bite, pile, amount	
Substance	all stuff without boundary or fixed shape, considered from a conceptual point of view not from a linguistic point of view; e.g. mass, material	Disjoint=Object
Symbol	anything used for conveying a message; e.g. traffic	



	sign, word, point, music	
Type	classes, races, kinds and sorts of entities; e.g. type, race, form, class, sort, series, version, release	

Functional is a class for many concepts that can apply to a diverse range of things (often unspecified) considered from a particular (pragmatic) conceptualization or perspective, e.g. “belongings”, “threat”, “cause”, “product”, “support”. Very often it cannot be classified as an Object or a Substance. The only way of specifying these concepts is by specifying the ROLE/INVOLVEMENT relation with a verb or noun denoting the conceptualization (or a BE\_IN\_STATE relation with a property or state). The class Functional just groups such vague meanings together. In some cases it is even unclear whether the involved entity is a FirstOrderEntity or a HighOrderEntity, so that it may be necessary to lift it to the highest level in the future. Other less obvious categories are Symbol and Type. Symbol represents all concrete things that have a meaning or are used for communication purposes. Type is only used for a small set of words that indicate types or kinds. Finally, the concepts Group and Part can be combined with other classes.

The class HighOrderEntity is subdivided into 6 main classes of which only the class Dynamic and Static form disjoint classes:

Dynamic	events or processes implying a change from one state to another state during the implied; event, act, action, become, happen, take place
Manner	the way in which an event takes place; e.g. manner, sloppy, strongly, way
Perception	mental or physical perception; notice, perceive
Phenomenon	any state or process that occurs in nature controlled or uncontrolled; e.g. weather, chance, luck, electricity
Static	all situations (properties, relations and states) in which there is no change: non-dynamic; e.g. state, property, be
Time	the duration or time point of a state or event; yesterday, day, pass, long, period

Only Dynamic and Static reflect a basic Aktions-Art distinction. We have encountered many problems with applying even these ‘fundamental’ distinctions to more specific classes of meanings. While general meanings could be classified either as dynamic or static, below these meanings there may be sub-fields for which the distinction is no longer clear. In all these cases the meanings have been lifted to higher Top-Concepts and each individual meaning has been classified separately with an additional link to Dynamic or Static. For example, we lifted the class of perceptions to a higher level Perception making it a co-hyponym of Dynamic and Static. Those meanings that are clearly dynamic perceptions can then have multiple hyperonyms to Perception and to Dynamic, clear static perceptions to both Perception and Static and unclear cases only to Perception:

Meaning	Relations	Top-Concept
dynamic perceptions: <i>look, hear</i>	HAS_HYPERONYM	Perception
	HAS_HYPERONYM	Dynamic
perception verbs: <i>itch</i>	HAS_HYPERONYM	Perception
static perception verbs: <i>pain</i>	HAS_HYPERONYM	Perception
	HAS_HYPERONYM	Static

In general, if it was difficult to verify a particular distinction at a more specific level then the specific level was promoted to a higher level (which was then created as a top-concept) and we have indicated that it can optionally be combined with other classes for those cases that yield clear intuitions (using multiple hyperonyms), whereas it can be under-specified when the intuitions are unclear. By using multiple hyponymy relations in this way the notions Top-Concepts are more and more being used as features.

The Dynamic concept is basically divided into three major classes: Activity, Event and Change. Change is used for acts and changes in which a single, specific state is continuously being changed during the event. Typical cases are achievements and accomplishments, where there is a natural termination point. Activities are on the other hand dynamic events that may continue over time, possibly in separate iterations, during which a variety of changes may take place but without referring to a single change as such. Typical examples are “sport”, “work”, “education”, “science”. BCs denoting such complex activities will often have subevent relations with various acts and processes which are *Changes*.<sup>2</sup>

Activity	all complex and planned events with a particular purpose and frequency; e.g. activity, hobby, sport, education, work, performance, fight, love, caring, management
BehavioralEvent	performances of beings; e.g. do, behave, act
CareEvent	all caring activities; e.g. nursing
Change	a single change either an act or a process
CommunicativeEvent	all actions with the purpose of communication; e.g. speak, tell, listen, command, order, ask, state, statement, conversation, call
ConsumptionEvent	; e.g. eat, drink, digest
EducationEvent	all activities and acts aiming at the training and learning of skills (mental and physical).
ManagementEvent	all controlling activities; e.g. rule, govern, control, guide
MentalEvent	a mental state is changed; e.g. invent, remember, learn, think,

<sup>2</sup> Note that an Activity can also be linked to all the participants in such an event without necessarily working out the ROLE/INVOLVEMENT. Take for example a script such as “fighting”. From a global point of view it is not clear who is the Agent and who is the Patient. A word such as “fighter” can thus only be linked by a general ROLE/INVOLVEMENT relation to the complex event of “fighting”. Related to “fighting” there may be subevents such as “hit”, “kick” or “hurt” which may have more specific roles with a patient, agent, or instrument.

	consider
MotionEvent	something changes location, irrespective of the causation of the change; e.g. move, put, fall, drop, drag, glide, fill, pour, empty, take out, enter
OperationalEvent	all activities or events in which an instrument is used or operated; e.g. operate, fly, drive, run
SoundEvent	all audible events; e.g. song, bang, beep, rattle, snore

Other more specific fields such as Art, Work, Management, Care, Communication contain both Activities and Changes or meanings which cannot clearly be classified as such. For naming purposes we will give these the extension Event which is then neutral with respect to Activity and Change. Events are then lifted up as a class and individual BCs can still be combined with Activity or Change using multiple hyperonyms.

The class Static is directly subdivided into various concepts dealing with the kind of property that is involved. Special classes are perhaps Meaning and Relation. Meaning only refers to the non-physical content that is being communicated. The physical representations have been classified as Symbol below FirstOrderEntity. The notion Relation is needed for meanings such as “kinship” where two first-order-entities are related but where the relation is not a property or state of each of those independently:

ExistentialState	all states which deal with the existence of objects and substances; e.g. exist, be, be alive, life, live, death
LocationState	; e.g. level, distance, separation, course, track, way, path
Meaning	the interpretation or message conveyed by a symbol or performance; meaning, denotation, content, topic, story, message, interpretation
MentalState	all states that exist in mind, possibly as the result of a mental process or event; e.g. knowledge, understand, motivation, idea, fact, truth, know
ModalState	states about possible situations as opposed to actual situations; e.g. ability, power, force, strength
PhysicalState	all perceptual and measurable properties of first order entities; e.g. health
QualityState	e.g. deficiency
Relation	a state which applies to a pair of object, states, events or processes; e.g. relation, kinship
SocialState	all states relevant to a particular society; e.g. unemployment, poor, rich

Since we do not classify adjectives in this project, most of the BCs belonging to these classes are nouns. These nouns often are the names for classes of adjectives, e.g. “colour”, “weight”, “size”. For these BCs the top-concepts are relatively more important because they do not exhibit much hierarchical structure among them. Without the top-concept

information it is impossible to distinguish between for example mental states and physical states from the relations between the word-meanings.

In Appendix IV the full hierarchy is given of all the 87 Top-Concepts (TCs). Appendix V contains the clustering of the BCs in this hierarchy. The print-outs are generated using the Novell ConceptNet1.0 tool (Cuypers and Díez-Orzas 1996) in which the top-concepts and the BCs are imported. All BCs are linked in ConceptNet with a hyponymy link. This hyponymy link should be considered as a rough cluster relation. When the relations of the BCs are worked out, it may very well be that some of the BCs are synonyms of the top-concepts to which they are assigned; some of them are hyponyms or sub-hyponyms; and yet others should be linked via causal, subevent, involvement or meronymy relations. In a few cases BCs have been assigned to multiple clusters but in other cases we have just selected one classification where multiple classifications could apply. In version 1.0 of ConceptNet it is not possible to specify disjunction or conjunction of multiple links. Such a specification should be made when the BCs are further defined by the sites for the individual languages.

It is important to realize that the TopConcept hierarchy does not necessarily correspond with the language-internal hierarchies. Each language-internal structure has a different mapping with the top-ontology via the ILI-records to which they are linked as equivalences. For example there are no words in Dutch that correspond with technical notions such as FirstOrderEntity, HighOrderEntity, but also not with more down-to-earth concepts such as Container. These levels will thus not be present in the Dutch hierarchy. From the Dutch hierarchy it will hence not be possible to simply extract all the *containers* because no Dutch word meaning is used to group or classify them. Nevertheless, the Dutch ‘containers’ may still be found either via the equivalence relations with English ‘containers’ which are stored below the sense of “container” or via the TopConcept clustering that is imposed on the Dutch hierarchy.

## 4. First Building Phase AMS

### 4.1 Tools and resources

AMS uses an object-oriented lexical database system (Boersma 1996) developed at the Computer Centrum Letteren of the University of Amsterdam for the Sift-project (LRE 62030). Some important characteristics of this lexical database are:

1. the object-oriented treatment of the data which makes it possible to efficiently manipulate lexicons, collections of entries, collections of senses or single entries and/or senses.
2. query any piece of information specified in a loaded dictionary
3. traverse explicitly coded semantic relations between word senses:
  - 3.1 in batch getting chains of senses or hierarchies at any depth
  - 3.2 via a so-called Surf interface, which is a hyper text representation of an entry in which links to other entries can be activated by clicking. These links can either be links within the same resource (to traverse semantic links from senses to sense) or across different lexical databases (e.g. going from a Dutch monolingual database to WordNet1.5 via a Dutch-English bi-dictionary).
4. edit information using:
  - 4.1 a general purpose editor in which any piece of information of an extracted entry or sense can be changed.
  - 4.2 a Surf editor in which specific fields, (especially the EWN relations) of an entry can be changed from the menus in the Surf-window.
5. compare, merge, insert and replace entries, senses or collections of various data types.

Within the AMS LDB we have loaded the following resources relevant for EuroWordNet:

- Celex Dutch lemma lexicon with basic syntactic information and corpus frequency information
- WordNet 1.5
- Vlis: the content of a lexical database provided by Van Dale.
- The Van Dale Dutch-English dictionary (Martin and Tops 1986).

In the near future also the Van Dale English-Dutch (Martin and Tops 1989) will be added.

In addition to this database there are several alternative tools and data available:

- definition parsers for English and Dutch definitions that provide the constituent structure of dictionary definitions (Vossen 1995).
- TRANS-CIRCLE: a program that groups synonymous meanings on the basis of overlapping occurrences as translations in two bi-lingual dictionaries.
- Extract Logical Form (ELF): a program that extracts the logical form of parse-trees representing constituent structures. The output is a list of triplet structures representing basic semantic relations (Vossen 1995, Vossen, Bon and Donker 1996).
- Aquilex Lexical Database (Carroll 1990) with several English and Dutch Machine Readable Dictionaries (Vossen 1995).

The data from the Van Dale Lexical Information System (VLIS) will be the input for developing the Dutch wordnet. The database contains the merge of several contemporary Dutch dictionaries published by Van Dale in recent years:

*Table 5: Number of entries and senses in VLIS database (AMS)*

	<b>nouns</b>	<b>verbs</b>
<b>entries</b>	63962	8822
<b>senses</b>	74678	14268

In this database particular semantic and morphological relations are explicitly coded for nouns, verbs and adjectives at sense level. Important semantic relations in VLIS are hyp(er)onymy, synonymy, antonymy, partitive and associative. Hyp(er)onymy is a ‘kind of’ relation from hyponym to hyperonym and vice versa. The difference in meaning is formulated in a field called ‘differentiae’. Whereas WN1.5 has applied multiple hyperonyms this was not allowed in the VLIS database.

Synonymy has been assigned (within part-of-speech) if there was no significant difference in meaning. With respect to synonymy VLIS has assigned the term ‘central’ to one sense and the term ‘peripheral’ to the other synonym(s). ‘Central’ denotes the most ‘general’ or frequently used term in Dutch. Hyp(er)onymy relations can only go from or to the central term and not to the peripherals.

Antonymy is defined as not X but Y and is only assigned to verbs and adjectives, e.g. ‘to succeed’ is an antonym of ‘to fail’. Partitive relations are assigned between nouns. The relation is used as the meronymy relation in WN1.5 but in VLIS it is not differentiated. Associative is always an additional relation between two senses that are closely related and possible between all categories. The idea was to assign this relation if another relation did not suffice. Therefore associative relations often stand for multiple hyperonymy, but also for near-synonymy or cross-part-of-speech synonymy.

Furthermore, there are other interesting semantic relations in VLIS, like causitive and inchoative but these are not assigned significantly. Finally, there is a morphological relation ‘reference’ of which it is unclear how it can be used semantically. The remaining

relations are either assigned insignificantly or variants to the synonymy or hyponymy relation.

In addition to the relations each sense contains a definition, some minimal syntactic properties, and it may contain labels with various types of information like: domains, attitude and style. We have extended these data with corpus-frequency information extracted from the Celex database.

#### 4.2 AMS approach

After evaluating the relations already present in the Vlis database we decided to take over particular relations, indicated in the next table:

Table 6: Conversion of VLIS relations (AMS)

	<b>Noun Senses</b>	<b>Verb Senses</b>	<b>Conversion to DutchWordNet</b>
<b>ABBREVIATION</b>	63	0	synset member
<b>ANTONYM</b>	181	148	NEAR_ANTONYM
<b>ASSOCIATIVE</b>	1763	557	--
<b>CAUSATIVE</b>	0	5	CAUSES
<b>HYPERONYM</b>	49815	8231	HAS_HYPERONYM
<b>INCHOATIVE</b>	0	11	CAUSES
<b>PARTITIVE</b>	576	0	HAS_MERONYM
<b>REFERENCE</b>	715	93	--
<b>SYNONYM</b>	19901	5049	synset member
<b>WOMAN/MAN</b>	95	0	synset member
<b>PREFERENCE</b>	2361	108	synset member
<b>FORM VARIANT</b>	4	0	synset member
<b>VERB</b>	0	474	HAS_HYPERONYM

In the VLIS database all relations are either from (target = false) or to (target = true) another sense. The number of relations counted are the relations with target = true. Some relations are meant to be bi-directional like antonymy and associative.

As will be clear from this table some EWN relations are missing and some are only incidentally coded. Especially, the work on the Base-Concepts (BCs) will be done manually as was agreed as a general policy. This will mean that the copied Vlis-relations are verified and the missing relations are added. Once the relations for the BCs are well-established we will use automatic techniques to extract additional information from the definitions. The extracted information will be used to:

- verify the copied Vlis relations. If the extracted information deviates from the copied information the relations will be verified manually.

- add those EWN relations which are not covered in Vlis

In addition to the extraction of information from the definitions we will use the bilingual dictionaries to (semi-)automatically extract equivalence relations with WordNet1.5 synsets and to extract potential new synset members. For the former we will implement the heuristics developed by FUE (D005) in our database. For the latter we will implement the TRANS-CIRCLE program in the AMS LDB.

For the manual coding we have developed a special Surf-Editor within the hypertext Surf windows of the AMS LDB. Using this editor it is not only possible to quickly browse or jump from entry to entry and sense to sense, but also to add, edit and remove relations. These relations can be specified between the selected sense and the sense activated in another Surf window. Using this option we can specify both language-internal relations and equivalence relations defined in EuroWordNet. The language-internal relations are automatically reversed, where relation features can be specified at both sides. The result of a specified relation is directly visual in the Surf-window and can be traversed.

Our experience so far with the manual coding is that it is very difficult, subjective and time-consuming to determine the exact synonymy and hyponymy of many closely related words, especially for the polysemous and vague BCs. Although the developed tests and criteria work for 'more distant' meanings they are not specific enough to always clarify the difference or the equality on a fine-grained level. Closely related to this problem are the difficulties to determine the appropriate level of polysemy of these words and, additionally, the matching to WordNet1.5 equivalences (see the general coding problems discussed in section 2). We have therefore chosen to give priority to the other relations for the BCs even though these may be less fundamental at first sight. Relations such as cross-part-of-speech equivalence, meronymy, subevent, cause and role elicit much stronger intuitions. Once these relations are established (as far as relevant) and after providing at least the correct hyperonym there is a better basis for determining the equivalence of closely related meanings (within a language and across languages).

We are currently processing the BCs in three cycles. In the first round (which has been finished) we make sure that all hyperonyms of the BCs are roughly ok and that we get a unified tree. Where possible other relations are added. In the second phase we have been storing more additional relations, we have included a full level of hyponyms for the BCs and we have covered some extensions. When necessary the hyperonyms have been adapted. This phase is still going on and has been completed for 25% of the BCs and some extensions. In a third phase we will process all the BCs again where we will specifically look at:



- the precise structure of the synset
- more precise hyponymy relations of closely related concepts
- the precise equivalence relation with WordNet1.5 synsets

The explicit structure given by the other relations (as far as they are present) will then give us better clues to determine the more exact synonymy, hyponymy and equivalence relations of very close meanings. If a set of word meanings are equivalent or have particular hyponymy relations then all the other specified relations should apply to the whole set.<sup>3</sup> For example, we have encountered a lot of parallelism between the hierarchies of verbs and nouns denoting the same event. In addition to a verb such as “handelen” (to merchandise) there is cross-part-of-speech equivalent noun “handel” (merchandising). Whereas “handelen” (to merchandise) has subevent-relations with “verkopen” (sell) and “kopen” (buy), “handel” (merchandising) has the same subevent-relation with the cross-part-of-speech equivalents “verkoop” (a selling) and “koop” (a buying). The cross-part-of-speech equivalence relations between these nouns and verbs can thus be exploited to create parallel relations as well. In this way the more exotic relation help determining the more fundamental relations.

### 4.3 AMS BC selection

As explained in D005 and D006 the Dutch BCs have been selected on the basis of four criteria:

- the number of relations
- the position in the hierarchy
- frequency in a corpus

where the latter criterion is only used in combination of one of the other criteria. According to section 3 the selected set overlaps with the common BCs in the following way:

Table 7: Dutch Initial Base Concepts, measured in WordNet1.5 senses (not synsets).

Site	POS	Total of AMS WN senses	Drop- Outs	Missing	Selected
AMS	Nouns	2265	1441	175	824
	Verbs	1174	737	34	437

The missing BC synsets for AMS have been added by looking for Dutch equivalences. This resulted in the addition of 94 new Dutch noun senses and 12 new Dutch verb senses.

<sup>3</sup> Equivalence is taken here as a broader notion, including genuine synonymy for synset members, near-synonymy and cross-part-of-speech synonymy.

Next we converted the set of common BCs back to Dutch senses. Since there is often more than one equivalent WordNet1.5 synset for a single Dutch sense this gives a much smaller set of senses.

*Table 8: Dutch Initial Base Concepts, measured in VlisSenses:*

Site	POS	Total of Vlis senses	Drop-Outs	Missing	Selected
AMS	Nouns	976	546	94	430
	Verbs	301	165	12	136

We can differentiate the drop-outs into the following subclasses:

- drop-outs that could be linked as (sub-)hyponyms of selected concepts using the Vlis hyponymy relations.
- drop-outs that have been linked to multiple wordnet Synsets of which one was selected by at least one other site and the other was only selected by AMS.
- remainder: genuine drop-outs.

The distribution of these classes is given in the next table:

*Table 9: Dutch Drop outs, measured in VlisSenses*

	<i>Nouns</i>		<i>Verbs</i>	
<b>Drop Outs</b>	535		160	
<b>(Sub-)Hyponyms</b>	358	65.5%	63	39.3%
<b>Multiple Equivalences</b>	34	6.0%	17	10.3%
<b>Genuine Drop Outs</b>	154	28.7%	80	50.0%

The drop-outs occurring at specific levels of selected senses have the following distribution over levels in the Dutch hierarchy:

*Table 10: Dutch Drop-Outs as Hyponyms of Selected BCs, measured in VlisSenses*

	<i>Nouns</i>	<i>Verbs</i>
<b>Total</b>	358	63
<b>Level 1</b>	283	54
<b>Level 2</b>	64	6
<b>Level 3+</b>	11	3

Remarkably, the drop-out verbs are to a much lesser extent located below the selected verbs. This is due to the fact that the verb hierarchy is relatively broad and less deep. It also means that the selection of Base Concepts has to be broader than it is now to capture the full diversity of the verb vocabulary. With respect to these classes we can say that the hyponym drop-outs do not represent new base concepts and they can be left out. In the case of the other drop-outs we can say that those with multiple translations are represented. The Dutch senses are already included by the other WordNet1.5 equivalences

that have been selected by at least one other site. The fact that AMS selected multiple senses however suggests that the non-selected WN-synsets should be added as well to the common BC selection. This has however no further consequences.

In the case of the remaining drop-outs (234 Dutch senses), we have considered the consequences of their classification for the top-ontology. The ontology should at least be capable of reflecting all the implied distinctions. In so far these drop-outs are more of the same it is not necessary to add them to the common set of BCs, in so far they represent new concepts which could not be represented in a proper way by the already selected concepts, they should be added and the other sites have to consider the relevance for their vocabulary. The latter was the case for 104 senses out of the 234 senses. In Appendix VI a full list is given of these drop-outs followed by its equivalences in WordNet1.5. The list is ordered by the TopConcepts to which they have been linked. Remarkably, the number of concrete drop-outs is very low. Most of the genuine drop-outs denote some HighOrderEntity. This shows that the selection of HighOrder-concepts is much less representative than the selection of the FirstOrder concepts, which was already suggested above and is not surprising when considering that the hierarchies of events, processes and states are more shallow. Since the total of genuine drop-outs is not very high we have classified all the 234 drop-outs with a TopConcept and added them to the Dutch BC-selection that is used in the first building phase.

Finally, we have included all the hyperonyms of the above senses to the subsets which are not within the set so far. In that way we guarantee that all concepts needed to define the selection are present. This gives then the following total set of 1049 Dutch senses that have been considered as the starting point:

*Table 11: Final Set of Base Concepts measured in Vlis senses (AMS)*

<b>Dutch BCs</b>	<b>Nouns</b>	<b>Verbs</b>	<b>Total</b>
<b>Initially</b>	976	301	1277
<b>Selected</b>	430	136	566
<b>Missing</b>	94	12	106
<b>Drop-Outs</b>	546	165	711
<b>(Sub)-Hyponyms</b>	358	63	421
<b>Multi-equivalences</b>	34	17	51
<b>Genuine Drop-Outs</b>	154	80	234
<b>SubTotal of Base Concepts</b>	678	228	906
<b>Missing Hyperonyms</b>	120	23	143
<b>Base-Concepts and their Hyperonyms</b>	798	251	1049

#### 4.4 AMS Results for the first subset

When coding the relations we have looked at the following issues:

- verify the copied hyperonym or add a missing hyperonym
- verify the adequacy of the top-concept link if present
- code any other EWN relation that was felt to be relevant for the meaning
  - specify cross-part-of-speech equivalences
  - specify meronymy, subevents, be\_in\_state, cause and role relations
- verify the quality of the synset
- verify the hyponyms
- verify the adequacy of the equivalence relations if present

As additional sources of information we have made use of:

- definitions
- the test-sentences developed for EWN
- human intuition
- morphologic structure of the entry
- WordNet1.5 classifications
- bilingual dictionaries

Initially, the coding process has been carried out for the 906 Dutch BCs and their 143 not-included hyperonyms, going from Top-Concept cluster to Top-Concept cluster. The first-order-entities have been processed once completely, the higher-order entities have been processed once completely and for 50% in a second revision. In addition to the BCs we have been working on some of the hyponym-extensions, especially movement-verbs, substance, food, and physical changes. These fields have not been covered completely in full depth, and the equivalence relations with WordNet1.5 have not yet been added. Furthermore, the BCs have been extended with concepts related via other relations such as meronymy, roles, causes, subevents, and cross-part-of-speech relations. In total this has lead to the addition of 2504 noun synsets and 1870 verb synsets. The total set of noun and verb senses covered by the 1<sup>st</sup> of February 1997 is given in table 12, where the BCs with their hyperonyms are presented as the *Core* and the last two columns give the figures for the Core and the Extension.

Table12: First Subset differentiated for Core and Core + 1<sup>st</sup>. Extension (AMS)

	<b>Core</b>		<b>Total 1st. Subset (Core + 1st. Extension)</b>	
	1049 synsets 4907 senses		5235 synsets 19240 senses	
	<b>Nouns</b>	<b>Verbs</b>	<b>Nouns</b>	<b>Verbs</b>
<b>SYNSETS</b>	798	251	3133	2102
number of senses (variants)	3342	1565	11729	7511
X variants per synset	4.6	6.7	3.7	3.5
corresponding to number of entries (words)	1661	731	5760	3130

<b>Equivalence to ILI (WN1.5)</b>				
<b>EQ_SYNONYM</b>	465	160	564	222
<b>EQ_NEAR_SYNONYM</b>	434	130	773	168
<b>HAS_EQ_HYPERONYM</b>	0	0	45	12
<b>total</b>	899	290	1382	402

<b>Top Concept Links</b>	569	208	855	449
--------------------------	-----	-----	-----	-----

The first striking aspect of this table is the ratio of variants per synset in the Dutch wordnet. For the Base Concept nouns and verbs the average is 4.5 and 6.7 variants per synsets, respectively. For the total first subset the ratios have dropped to 3.7 and 3.5 respectively. Still this is considerably more than in WordNet1.5, where the average number of synset members or variants is 1.8 (Díez-Orzas 1996). However, here we are here mostly dealing with general vocabulary. When we extend the first subset we expect that the average in Vlis will decrease. If we take the synonymy-relations for Vlis as a whole the average number of synset-members is 1.33, which is lower than WordNet1.5. Also note that we have not focused on the synonymy relations of the senses in the current phases. Changes therefore mainly involved very obvious omissions or senses that could not be distinguished. We expect that the synset-structure will be changed much more in the next phase and after comparing the results with the other wordnets.

The table further shows that most Dutch Base Concepts (960 senses) have both a Top Concept and at least one WordNet1.5 equivalent. Most of the equivalence specifications still come from the translation of the BCs, before the actual coding has started. We have postponed the work on the equivalence relations for several reasons:

- the necessary resources have only recently been loaded in the AMS LDB.
- we think that it is easier to specify equivalence relations after we have a clear picture of the monolingual structure.

- future comparison of the monolingual structure and the possible measures to make the cross-linguistic mapping more efficient will determine the final strategy for specifying the equivalence relations.
- the automatic matching heuristics designed by FUE will be implemented in the AMS LDB as well.

Some of the given equivalence relations have been verified and sometimes adjusted. The kind of equivalence relation has not yet been differentiated. Certainly when a Dutch sense is linked to multiple WordNet synsets or multiple Dutch senses to a single WordNet synset we will have to differentiate the equivalence relation into NEAR\_EQ\_SYNONYM.

In total 1,304 Dutch senses have an explicit link with a TopConcept. In most cases the link is NEAR\_SYNONYM. As explained above the link should be seen as a global cluster link. The Dutch meanings could actually be a hyponym or a sub-hyponym of the TopConcept. The coverage is 25% of the total set of 5,235 synsets considered. However, most of the covered senses are directly or indirectly linked to another Dutch sense which has such a link. At first the TopConcept links have been taken over from the BC-clustering applied to the WordNet1.5 synsets, via the translation relation. The ConceptNet1.0 hierarchy of the ontology is used for verification of the clustering of the Dutch senses (see Appendix IV). In addition to the BCs we have assigned a TopConcept link to many verbs and nouns that form very shallow hierarchies. Especially in the case of *changes* and *states*, hierarchies are extremely flat, making it impossible to get coherent groupings. There is no differentiation between for example physical changes and mental changes or between modal states or physical states. All these nouns and verbs can only be linked as a hyponym of the equivalent of *change* and *state*.

Relatedness of meaning only follows from the kind of property that is involved. In the case of changes this can be expressed by a CAUSES relation to the resulting state, in the case of properties and states by a cross-part-of-speech relation. However, such an adjective does not always exist and the adjectives as such are not classified or linked in EuroWordNet (due to limitations in the budget). Consequently we cannot infer from the related adjective (if present at all) that we are for example dealing with a physical process or not. Therefore, we have specified a top-concept for all these nouns and verbs as well:

Dutch Sense	Relation	Target Concept
<b>stollen V</b> (to become solid)	<b>HAS_HYPERONYM</b>	<b>veranderen V</b> (to change)
	<b>CAUSES</b>	<b>vast A (solid)</b>
<b>afstompen V</b> (to become insensitive)	<b>BELONGS TO TC</b>	<b>PhysicalChange</b>
	<b>HAS_HYPERONYM</b>	<b>veranderen V</b> (to change)
<b>gewicht N</b> (weight)	<b>CAUSES</b>	<b>afgestompt A</b> (insensitive)
	<b>BELONGS TO TC</b>	<b>MentalEvent</b>
	<b>HAS_HYPERONYM</b>	<b>eigenschap N</b> (property)
	<b>XPOS_NEAR_SYNONYM</b>	<b>wegen V</b> (to weight)
	<b>XPOS_HAS_HYPONYM</b>	<b>zwaar A</b> (heavy)
	<b>BELONGS TO TC</b>	<b>PhysicalState</b>

These Top-Concept links can thus indirectly be seen as a global classification of the adjectives or states incorporated in the meanings of nouns and verbs.

The next table gives the full list of language-internal relations coded for both the Core of Base Concepts and the Extension. Obviously, hyponymy is the most-frequently coded relation in both the Core and the Core+Extension. The average number of language-internal relations or density is 3 per BC-noun and 4 per BC-verb, which drops to 2.7 for the extended set. If we exclude the hyponymy-relations we see that the density for BC-nouns is still 0.9 and for BC-verbs even 1.9. This means that at average every noun has one non-hyponymy relation and every verb two. This is in line with the followed strategy to provide a rich encoding of all the relevant relations for the BCs.

Table13: Relations Encoded for First Subset (AMS)

Language Internal Relation	Core		Total 1st. Subset (Core + 1st. Extension)	
	Nouns	Verbs	Nouns	Verbs
BE_IN_STATE	10	0	22	0
CAUSES	27	89	56	420
HAS_HYPERONYM	791	252	3316	2194
HAS_HYPONYM	791	252	3316	2194
HAS_HOLONYM	10	0	33	0
HAS_HOLO_LOCATION	2	0	18	0
HAS_HOLO_MADEOF	5	0	17	0
HAS_HOLO_MEMBER	11	0	54	0
HAS_HOLO_PART	39	0	146	0
HAS_HOLO_PORTION	8	0	34	0
HAS_MERONYM	73	0	187	0
HAS_MERO_LOCATION	0	0	18	0
HAS_MERO_MADEOF	6	0	17	0
HAS_MERO_MEMBER	27	0	55	0
HAS_MERO_PART	51	0	146	0
HAS_MERO_PORTION	11	0	35	0
HAS_SUBEVENT	29	30	41	65
HAS_XPOS_HYPERONYM	2	3	4	12
HAS_XPOS_HYPONYM	4	3	17	4
INVOLVED	0	5	2	43
INVOLVED_AGENT	0	3	3	14
INVOLVED_DIRECTION	0	0	0	1
INVOLVED_INSTRUMENT	2	7	3	76
INVOLVED_LOCATION	1	3	1	7
INVOLVED_PATIENT	8	28	27	116
INVOLVED_TARGET_DIRECTION	0	3	0	10
IS_CAUSED_BY	50	71	89	124
IS_SUBEVENT_OF	14	32	30	76
NEAR_ANTONYM	26	54	66	118
NEAR_SYNONYM	54	20	99	29
ROLE	15	0	51	1
ROLE_AGENT	15	0	19	0
ROLE_DIRECTION	0	0	2	0
ROLE_INSTRUMENT	13	0	79	0
ROLE_LOCATION	4	0	8	0
ROLE_PATIENT	77	0	144	2
ROLE_TARGET_DIRECTION	2	0	10	0
STATE_OF	1	0	3	2
XPOS_NEAR_ANTONYM	3	3	3	3
XPOS_NEAR_SYNONYM	72	86	205	198
Total	2254	944	8376	5709



A problem for describing the actual work in terms of the language-internal relations is the fact that particular relations have been taken over from the Vlis database. In many cases we have not changed the relations from Vlis, in other cases we have added relations, removed relations or we have edited or modified a relation between two specified senses. To overcome this problem we have added a so-called status field which specified whether the relation was copied, manually added, revised and whether a relation is considered as OK. The latter status tag is only assigned when we verify the relations in the second phase. Once the coding has been completed we can take the senses with all the relations that have this label assigned. Currently, the data for the first subset is only partially coded as such.

The following tables then show the EuroWordNet relations again where we have indicated how many relations have been copied originally (the second column), how many relations have been maintained after verifications (column 3) and how many relations have been added manually (the 4<sup>th</sup> column). The tables have been subdivided for nouns and verbs and for the Core and the full subset. For the BC-nouns we see that 35% of the total number of relations was taken over and 65% was added manually. The copied relations mainly involved the hyponymy and antonymy relations, where the meronyms have mostly been differentiated into subtypes. However, even in the case of the hyperonyms we see that almost 50% has been added manually. These manually added hyperonyms fall into three classes:

- corrections when the wrong hyperonyms was assigned.
- cases when there was no hyperonym: the tops in the Van Dale database.
- cases where another hyperonym applied as well.

Table 15 for the BC-verbs shows an even more extreme portion of manual addition: 81%. This has two reasons:

- as we saw before more non-hyponymy relations have been added for verbs than for nouns
- Van Dale has devoted less time on coding the verbs, consequently many verbs still have no hyperonym

Typically, we see that only 105 hyperonyms were specified for 232 verb synsets, less than 50% has a hyperonym.

Table 14: Copied and added relations for 798 BC-Nouns (AMS)

<i>Relations</i>	<i>Originally Copied</i>	<i>Confirmed Copied</i>	<i>Manually Added</i>	<i>Total</i>
BE_IN_STATE			10	10
CAUSES			27	27
HAS_HOLONYM	7	3	7	10
HAS_HOLO_LOCATION			2	2
HAS_HOLO_MADEOF			5	5
HAS_HOLO_MEMBER		1	10	11
HAS_HOLO_PART		2	37	39
HAS_HOLO_PORTION		1	7	8
HAS_HYPERONYM	485	400	391	791
HAS_MERONYM	96	66	7	73
HAS_MERO_MADEOF			6	6
HAS_MERO_MEMBER		2	25	27
HAS_MERO_PART		3	48	51
HAS_MERO_PORTION		1	10	11
HAS_SUBEVENT		1	28	29
HAS_XPOS_HYPERONYM			2	2
HAS_XPOS_HYPONYM			4	4
INVOLVED_INSTRUMENT			2	2
INVOLVED_LOCATION			1	1
INVOLVED_PATIENT			8	8
IS_CAUSED_BY			50	50
IS_SUBEVENT_OF			14	14
NEAR_ANTONYM	26	19	7	26
NEAR_SYNONYM			54	54
ROLE			15	15
ROLE_AGENT			15	15
ROLE_INSTRUMENT			13	13
ROLE_LOCATION			4	4
ROLE_PATIENT			77	77
ROLE_TARGET_DIRECTION			2	2
STATE_OF			1	1
XPOS_NEAR_ANTONYM			3	3
XPOS_NEAR_SYNONYM			72	72
Total	614	499	964	1463
Percentages		35%	65%	

Table 15: Copied and added relations for 251 BC-Verbs (AMS).

<i>Relations</i>	<i>Originally Copied</i>	<i>Confirmed Copied</i>	<i>Manually Added</i>	<i>Total</i>
CAUSES	1	2	87	89
HAS_HYPERONYM	105	83	169	252
HAS_SUBEVENT			30	30
HAS_XPOS_HYPERONYM			3	3
HAS_XPOS_HYPONYM			3	3
INVOLVED			5	5
INVOLVED_AGENT			3	3
INVOLVED_INSTRUMENT			7	7
INVOLVED_LOCATION			3	3
INVOLVED_PATIENT			28	28
INVOLVED_TARGET_DIRECTION			3	3
IS_CAUSED_BY	5	8	63	71
IS_SUBEVENT_OF			32	32
NEAR_ANTONYM	44	39	15	54
NEAR_SYNONYM			20	20
XPOS_NEAR_ANTONYM			3	3
XPOS_NEAR_SYNONYM			86	86
Total	155	132	560	692
Percentages		19%	81%	

If we look at the total subset we see that the proportion of manual additions decreases for both verbs and nouns:

Table 16: Copied and added relations for total First Subset Nouns and Verbs (AMS).

	<b>Originally Copied</b>	<b>Confirmed Copied</b>	<b>Manually Added</b>	<b>Total</b>
Number of relations for nouns	3236	2489	2575	5064
Percentages for nouns		50%	50%	
Number of relations for verbs	2007	1723	1792	3515
Percentages for verbs		49%	51%	

Especially for the verbs the decrease is drastic. The main reason for this is that hyponymy is better coded for the more specific verbs (84% has a hyperonym), another reason is that the extension (both for nouns and verbs) for some extent only involved adding hyponyms. The non-hyponymy relations are only added for more specific concepts when they can be derived semi-automatically.

For completion, we give in the following table the relations coded to senses belonging to other parts-of-speech (mainly adverbs and adjectives).

Table 17: Copied and added relations for other Parts of Speech (AMS).

<i>Relations</i>	<i>Originally Copied</i>	<i>Confirmed Copied</i>	<i>Manually Added</i>	<i>Total</i>
HAS_HYPERONYM	1	1	0	2
HAS_XPOS_HYPERONYM	0	0	5	5
IS_CAUSED_BY	0	0	260	260
STATE_OF	0	0	16	16
XPOS_NEAR_SYNONYM	0	0	19	19
Total	1	1	300	302

Since we do not consider the relations among adjectives and adverbs themselves, here the causal and cross-part-of-speech relations dominate.

Finally we give an overview of the main semantic areas covered so far. The next table gives the meanings which have the most direct and indirect hyponyms in the limited set: i.e. only taking into account the meanings in the first subset of meanings):

Table 18: Major Semantic Fields of AMS First Subset in terms of hyponyms

54 slaan 1 (to give a blow)	55 beweging 1 (movement)	61 ervaren 2 (experience)	57 verschijnsel 1 (phenomenon)
58 houding 2 (attitude, position)	60 taaluiting 1 (language utterance)	69 voortbrengen 1 (to produce)	66 gedrag 1 (behaviour)
67 maken 2 (to make)	67 uiting 2 (utterance)	73 zijn 7 (to be)	70 nemen 1 (to take)
73 bezighouden 3 (to occupy)	73 meemaken 1 (to experience)	82 mens 1 (human)	76 daad 1 (deed)
76 persoon 1 (person)	76 tonen 2 (to show)	105 voedsel 1 (food)	84 maat 5 (measure)
91 raken 2 (to hit)	97 deel 2 (part)	114 wezen 1 (being)	107 organisme 2 (organism)
110 weergave 2 (representation)	113 voortbewegen 2 (to move)	123 geestesgesteldheid 1 (state of mind)	116 verwerven 1 (acquire)
119 vorm 1 (form)	121 middel 2 (means)	156 bezigheid 1 (activity)	134 uitoefenen 2 (to act upon)
140 gebruiken 1 (to use)	151 handeling 1 (action)	182 gaan 3 (to go)	158 gesteldheid 1 (constitution)
170 verandering 1 (a change)	180 groep 4 (group)	232 handelen 2 (to act)	185 bewegen 1 (to move)
209 verplaatsen 1 (to place)	220 voorwerp 1 (object)	382 eigenschap 1 (property)	307 stof 4 (substance)
314 toestand 1 (situation)	342 doen 5 (to act)	1015 veranderen 4 (to change)	440 veranderen 1 (to change)
599 gebeurtenis 1 (event)	670 bewegen 2 (to move)	1206 veroorzaken 1 (to cause)	2307 gebeuren 2 (to take place)
2819 iets 1 (anything)			

The meanings represent the major nodes after the restructuring. The most important concepts are “iets 1” (anything), “gebeuren 2” (to take place) and “zijn 7” (to be) because they are the tops of the full hierarchy. Below “iets 1” we find all first-order-entities but also the ‘dynamic’ noun “gebeurtenis 1” (event), and the static nouns “toestand 1” (situation), and “eigenschap 1” (property) which are top-meanings for high-order-entities. Below are some of the hierarchical dependencies for the higher-order-meanings:

iets 1 (anything)		
HAS_HYPONYM		gebeurtenis 1 (event)
HAS_XPOS_HYPONYM		gebeuren 2 (take place)
HAS_HYPONYM		toestand 1 (situation)
HAS_HYPONYM		eigenschap 1 (property)
HAS_XPOS_HYPONYM		zijn 7 (be)
HAS_HYPONYM		voorwerp 1 (object)
HAS_HYPONYM		stof 4 (substance)
toestand 1 (situation)		
NEAR_SYNONYM		eigenschap 1 (property)
XPOS_NEAR_SYNONYM		zijn 7 (be)
gebeuren 2 (take place)		
XPOS_NEAR_SYNONYM		gebeurtenis 1 (event)
HAS_HYPONYM		veranderen 1 (to change uncontrolled)
	HAS_HYPONYM	bewegen 1 (move)
	HAS_HYPONYM	krijgen 1 (receive)
	HAS_HYPONYM	veroorzaken 1 (to cause)
HAS_HYPONYM		veranderen 4 (to change controlled)
	HAS_HYPONYM	bewegen 2 (move)
	HAS_HYPONYM	bewerken 1 (to modify)
	HAS_HYPONYM	voorzien 3 (to provide)
	HAS_HYPONYM	maken 2 (to make)
HAS_HYPONYM		handelen 2 (to act)
HAS_HYPONYM		doen 5 (to do)
etc.....		
gebeurtenis 1 (event)		
XPOS_NEAR_SYNONYM		gebeuren 2 (take place)
HAS_HYPONYM		verandering 1 (to change)
	HAS_HYPONYM	beweging 1 (move)
HAS_HYPONYM		handeling 1 (an act)
HAS_HYPONYM		daad 1 (a deed)

This hierarchy illustrates the phenomenon discussed above that the verbal hierarchy to a large extent is mirrored in the higher-order noun hierarchy. This phenomenon is persistent at deeper levels.

#### 4.5 Spent resources and prospects for the extendibility

We think that the Van Dale lexical database provides a good starting point for developing a Dutch WordNet. Because of the availability of these basic relations we can focus on a rich and high-quality encoding of the relations of the Base Concepts. This is shown in the amount of manual work and the density of relations:

- 65% to 81% addition of relations
- high density of relations per sense: 3-4 language-internal relations average per synset

The resulting network of the most basic and important meanings can be seen as a solid foundation for the complete Dutch Wordnet. The resources spent for manually coding the above relations are estimated in the following table:

Table 19: Effort Spent for building the AMS First Subset with extensions

<i>Types of relations</i>	<i>Hours 7/day</i>	<i>Days (7 hours)</i>	<i>Months (22days)</i>	<i>No. Of Relations</i>
<b>Equivalence relations</b>	144	20.5	0,93	1784
<b>Top Concepts</b>	105	15	0,68	1214
<b>Language Internal Relations</b>	236	34	1,55	8575
<b>Total</b>	485	74	3,16	11573
<b>Effort spent for the concept “gebouw” (building)</b>				
<b>Equivalence relations</b>	4	94	23,5	
<b>Language Internal Relations</b>	12	525	44	

With manually coding we mean verifying the copied Vlis relations and making changes and/or additions. The last two rows of the table also show the effort for one extension: all words related to *building*. According to this the average speed is 23.5 synsets/hour. The remaining work is shown in table 20, where the ratio of synsets and synset-members or senses is expected to decrease for the remaining extension:

Table 20: Remaining Effort Spent for Extension of the concept “gebouw” (building)

	<i>Current coverage</i>	<i>Core</i>	<i>Final coverage</i>	<i>Remaining non-core</i>
<b>senses</b>	<b>20.000</b>	<b>4.907</b>	<b>50.000</b>	<b>30.000</b>
<b>synsets</b>	<b>5.235</b>	<b>960</b>	<b>25.000</b>	<b>20.000</b>

The total effort for the remaining work is shown in the next table, taking the above coding-speed as a basis:

Table 21: Estimated Effort for the remaining work (AMS).

<i>Types of relations</i>	<i>No. Of Relations</i>	<i>Synsets /hour</i>	<i>Effort Core (hours)</i>	<i>Days (7 hours)</i>	<i>Months (22days)</i>
<b>Language Internal Relations</b>	<b>20.000</b>	44	454	65	3
<b>Equivalence relations</b>	<b>20.000</b>	23,5	851	121	5,5
<b>Total</b>					<b>8,8</b>

Note that the coding-speed was based on a manual process only. The estimated 8.8 person/months is more than available: 6 person/months. However, the full extension is still feasible because of the use of semi-automatic techniques, which are more reliable and efficient for more specific meanings. Furthermore, we do not intend to cover all the relations for the extension. We will mainly focus on hyponymy and synonymy and only extract other relations when there are clear patterns that can be extracted automatically.

## 5. First Building Phase PSA

We consider this first phase of work, dedicated to the initial construction of the Italian wordnet and the linking of the Italian base concepts to the Interlingual Database (currently identified in WordNet1.5), as a fundamental stage in the project life-time. Until now, together with the other partners, we have worked i) at defining the structure and the main semantic relations to be represented in our concept nets; ii) at identifying the base concepts for the languages being treated and the ways in which cross-language equivalences can be recognised and coded. In the first building phase, we are now beginning to test our original hypotheses, formulated on the basis of subsets of data from our sources, over the totality (or at least very large subsets) of our Italian lexical data. We thus have the chance to evaluate the consistency of our sources, to test how far the various types of semantic relations can be extracted (semi-)automatically from them, to recognise any previously unforeseen difficulties and to take the necessary steps to overcome them. A major question is to what extent are our sources, built in the first place by lexicographers and then at least partially analysed and restructured by linguists, suitable for the construction of a new tool: a semantic concept net. I.e., how much and what kind of work needs to be done on the data; to what extent can we operate automatically and what manual interventions are necessary; how time consuming is all this?

### 5.1 Tools and Resources

#### 5.1.1 MRDs and Lexical Databases as input

At Pisa, we have three main sources for the Italian data:

1. *The Italian Monolingual Lexical Database (constructed from a number of sources)*

The current figures for the LDB used for the EuroWordNet project are:

*Table 22: Entries, Senses and Examples in the Italian LDB*

	Verbs	Nouns	Total
No. of Lemmata	5,546	24,635	30,181
No. of Word senses	14,091	45,608	59,699
No. of Example Sentences	7,815	7,815	



Table 23: Taxonomic Data represented in the Italian LDB<sup>4</sup>

For Nouns:	Genus (IS-A + Syn)	Part_Of	Set_Of	V to N	A to N
Total No. of Occurrences	36,522	1,500	1,558	5,379	972
No. of Different Words	6,070	643	726	2,712	895

For Verbs:	Genus (IS-A + Syn)	Causative
Total No. of Occurrences	13,676	415
No. of Different Words	2,593	180

This monolingual LDB is the main source of data for the Italian wordnet; the semantic relations, with the exception of the synonym and antonym relations, are extracted (when present) from this source.

#### 2. Italian/English Bilingual Lexical Database (constructed on the basis of the Collins-Giunti Italian-English dictionary)

Approximately 30,000 senses on each side (Italian-English, English-Italian). This data is used to provide a first translation of the Italian word-senses and as a source of potential synonyms.

#### 3. An Electronic Dictionary of Synonyms available at ILC-CNR.

Used as a source for indications on synonym data and of additional word senses.

#### 4. Native Speaker Intuition

A fourth source of data which should perhaps be mentioned is that derived from our own intuition and linguistic competence (see Local Approach below).

### 5.1.2 Local Tools and Technology

Three separate sets of tools have been/are being used in the various stages of processing.

1. The general dictionary processing system used is the *Pi-System* developed by Eugenio Picchi, ILC-CNR. The main components of the *Pi-System* used in the preliminary analysis, structuring and extraction of the lexical source data for the construction of the Italian Wordnet are listed below. These tools are in different stages of implementation: some have been industrialised (e.g. the basic version of the DBT system) and are generally available, others are available as prototypes for research purposes, others are now being tested in the Institute.

<sup>4</sup> It is necessary to clarify that with regard to the verbs we did not have disambiguated data. This means that we found the indication of an IS-A or synonymical or 'Causative' relation between a verb sense and a verb homograph, not a verb sense (cf. D006), and we are manually disambiguating all our data .

- *DBT*: system to create, manage, interrogate text databases (structured and unstructured text)
  - *Morphological Engine* (currently running for Italian and English)
  - *MLDB*: system for acquisition, management, querying of mono- and bilingual Lexical Databases (LDBs)
  - *MISC*: semi-automatic sense merging of information from different LDBs
  - *LEXICOGRAPHIC WORKSTATION*: integrated procedure for dictionary creating and editing (includes interactive sense disambiguation procedure).
2. The tool which has been used for the extraction of part of semantic data from the dictionary entries is the *AcqSys-System for the Acquisition of Taxonomic Information from Dictionary Definitions* developed by Simonetta Montemagni. This tool adopts a two stage approach:
- i. initial syntactic analysis for each definition
  - ii. pattern matching procedure operating on syntactic description to extract taxonomic data
- Stage (i) uses:
- general purpose Italian grammar + 2 modules: *p-Fit* - fitting procedure which attempts to solve failed analyses; *p-DAm* - attempts to solve disambiguities and malformations (typical of dictionary definitions)
- Stage (ii) uses:
- Genus extractor + other components for verb frame extraction (from examples).

The following relations are encoded:

For Nouns:

IS-A	(hyperonym)
SYN	(synonym)
TYPE_OF	(hyperonym)
SET_OF	(holonym)
PART_OF	(meronym)
ELEM_OF	(meronym)
V to N	(deverbal - morphologic and non)
A to N	(deadjectival - morph. and non)
AGENT_OF	(morphologic and non)

For Verbs (but cf. footnote 4):

IS-A	(hyperonym)
SYN	(synonym)
ANT	(antonym)
CAUS (fare+inf)	(causative genus)
CAUS (rendere+inf)	(causative genus)
INCH (cambiare/diventare)	(inchoative genus)

3. We are now developing a modular set of procedures to process the source data and map them onto the EWN import format. These procedures include modules which treat the source data, i.e. extract and map the relevant data onto a predefined input template, and modules which process the extracted data preparing it for wordnet building (e.g., procedures for consistency checking and word sense numbering and renumbering). These procedures form the core of the Italian Concept Net Building Set.

## ***5.2 Local approach***

At Pisa, in accordance with the general project objectives, we have two main aims: to construct a consistent basis for our wordnet which will represent not only generally recognised concepts but also the language-specific features of Italian and will prove to be a flexible and useful tool for various kinds of language processing activities. This means that we must respect three (apparently contradictory) maxims: our concept net must be rigorously coherent in its structure; cross-language compatibility must be ensured; it must be intuitively credible to the native speaker.

As stated in the Introduction, the data sources and the processing strategies to achieve the common goals differ from partner to partner. At Pisa we decided to construct the Italian wordnet from a number of sources (at least, at the upper level of the taxonomies) in order to be able to overcome, to some extent, the idiosyncracies of a single dictionary and provide a more objective perspective on the data. This approach has certainly required more work than if we had just considered a single source. Our starting point was thus the creation of the Base Concept synsets; these were constructed using data from 3 different sources (see below 5.3). However, an integration of different sources has also highlighted the differences between and the inconsistencies found in dictionary data; word senses and synonyms vary from source to source. This has meant the necessity of human intervention; the synsets are thus evaluated by several native language speakers with a high level of competence. Thus, in this first phase the building of the BC synsets has involved a considerable effort in Pisa. We have been willing to dedicate far more time to this stage than allowed for in the TA because it represents the foundation on which the whole structure of our concept net will be based.

## ***5.3 PSA BCs selection***

### ***5.3.1 Criteria and methods adopted for the preliminary selection***

As explained in D005 and D006, we first identified a preliminary set of BCs, by automatically extracting a list of lexical items from the Italian monolingual LDB (using data elaborated by *AcqSys-System* - see above), primarily in consideration of their position (medium/high) in the taxonomy and also of the number of relations with other lexical items (generally hyponyms). This first proposal was then processed manually and a preliminary

uniting of items expressing the same concept was made. This list of word senses/synsets was then mapped to WN1.5 in order to establish cross-language lexical equivalences. A number of problems were identified at this stage: it was difficult to identify the correct sense, to distinguish between close senses - often the choice was arbitrary or one Italian word-sense was mapped to more than one WN1.5 sense (these problems have been documented and exemplified in deliverables D005 and D006).

According to what explained in section 3, our preliminary selection was compared (by AMS) with those of the other sites and a common set of BC synsets was determined. In the following table the results of this comparison for the Italian BCs are shown:

Table 24: Italian Initial Base Concept Synsets (PSA)

	<b>Nouns</b>	<b>Verbs</b>
<b>Initially identified</b>	287	101
<b>Selected as common BCs (Union of intersection pairs)</b>	497	131
<b>Italian BCs selected</b>	190	47
<b>Missing from the original selection</b>	307	84
<b>Drop-outs</b>	93	54

### 5.3.2 Treating ‘Missing’ concepts and ‘Drop-outs’

We then took into consideration the set of base concepts selected by other partners but not by Pisa (the so-called ‘missing’ concepts). It was necessary to follow an inverse procedure to obtain cross-language equivalences. From the WN1.5 sense definitions, we had to select the most nearly equivalent Italian word senses. This was not easy and it was not always possible to find an Italian equivalent. At times, the WN1.5 sense given was so general that no Italian word seemed to quite fit. For example, one of our missing concepts was defined in WN1.5 as: *follower* 1 - an ordinary person who accepts the leadership of another. Strange as it may seem we were unable to identify an Italian term that has this general, unnotated sense of ‘follower’. The apparent equivalent *seguace* is defined as “chi segue e sostiene una dottrina, una scuola, un maestro, etc.” (who follows and supports a doctrine, a school, a master, etc.), where there is the implicit idea of ‘strong belief’ rather than the ‘acceptance’ implied by the WN1.5 definition. Actually, we had already included in our set of base concepts the synset {*seguace*, *discepolo*} which mapped to WN1.5 {*disciple*, *adherent*}, a hyponym of *follower* 1. This synset is thus also mapped to the missing base concept as a cross-language equivalent hyponym.

In some other cases, the English sense chosen lexicalises a concept which is not lexicalised in a single lexical unit in Italian. For instance, one verb concept not contained in the preliminary Italian selection was that of *to ride* which does not really exist in Italian as a single lexical item. Furthermore, we do not have any corresponding verb for its hyponyms either (partly reported under (i) below), while we have the corresponding hyperonyms ((ii) below):

i)

ride -- (be carried or travel on or in a vehicle; "I ride to work in a bus"; "He rides the subway downtown every day")

=> chariot -- (ride in a chariot)

=> bicycle, cycle, bike, pedal, wheel -- (ride a bicycle)

=> backpedal -- (pedal backwards, as on a bicycle)

=> motorcycle, cycle -- (ride a motorcycle)

=> train, go by train, ride a train, ride in a train -- (travel by train)

=> sled, sleigh -- (ride (on) a sled)

=> bobsled, ride a bobsled

=> toboggan, luge

..... etc. ....

ii)

ride -- (be carried or travel on or in a vehicle; "I ride to work in a bus"; "He rides the subway downtown everyday")

=> travel -- (undergo transportation, as in a vehicle)

=> travel, go, move, locomote -- (change location; move, travel, or proceed; "How fast does your new car go?" "We travelled from Rome to Naples by bus"; "The policemen went from door to door looking for the suspect").

In Italian it is extremely difficult to find a reference to 'motion' conflated with 'vehicles' in the meaning of a verb, and we can in no way identify lexical items corresponding to *ride* and its hyponyms; thus, we may only provide a translation for its hyperonyms.

In still other cases, we can find an Italian synset roughly corresponding to one missing from our original selection, however either its hyperonyms or its hyponyms are completely different from those found for the WN1.5 synset. For example, *{to neaten, to groom}* (defined in WN1.5 as "care for the external appearance") could match with *{rassettarsi, aggiustarsi, sistemarsi}*: however, if we take into consideration the Italian synset taxonomy, we see that while a correspondence with the English one can be found at the hyperonym level, there is no possibility to identify matchings at the level of the hyponyms. Therefore, we may only state an *eq\_near\_synonym* relation between the Italian and the English synsets (in any case, it must be remembered that we are still in a preliminary stage of our work and we are thus now experimenting solutions which need to be verified and which, in some cases, will be modified by further acquisitions and/or evaluations.)

When examining our list of missing concepts, at times we also found that we had a very similar sense in our list of 'drop-outs' (the base concepts which had been selected for Italian and not for the other languages) - maybe we had chosen to map it to a WN1.5 sense which was one level higher or lower in the hierarchy. In these cases, we re-examined our original proposal and, when possible, mapped this sense directly to the appropriate missing concept. In other cases, we encoded an appropriate equivalent Italian sense to the missing item but also maintained the drop-out (as being more representative of the Italian data).

For instance, a synset selected, but missing in the Italian selection, is that of {*supply, provide, render, furnish*}, which has the following hyperonym synsets:

supply, provide, render, furnish

=> give -- (transfer possession of something concrete or abstract to somebody; "I gave her my money"; "can you give me lessons?")

=> transfer -- (cause to change ownership)

One of our drop-outs is the {*dare*} synset, corresponding to {*give*}, which is not found as a hyperonym of the Italian synset for {*supply, provide, render, furnish*} in our source (but this seems more a shortcoming of our source). In this case, if we did not recover our drop-out synset we would lose a significant part of our data, since this synset has several hyponyms in Italian.

Thus, we had to carefully reconsider all our drop-outs, since they are major concepts in our language and play a central role in the definitions of many other verbs. By this analysis, we have seen that they can be differentiated into some sub-groups:

- i) drop-outs which turn out to be hyponyms/hyperonyms of the BCs falling in the union of intersection pairs, and so will be in any case included in the subsets which will be encoded;
- ii) drop-outs originally linked to different WN 1.5 synsets, due to the difficulties reported above in establishing equivalence relations; indeed, after reconsidering more carefully the synsets in the union of intersection pairs, we have seen that some of our drop-out BCs could as well be related to synsets selected by the others (i.e., these BCs could map to different WN1.5 synsets, due to the extreme sense-differentiation in WN 1.5);
- iii) 'real' drop-outs.

In general we have included our drop-out concepts in our BC list when (i) although corresponding very closely to one of the missing concepts, they are found at a different level in the Italian taxonomy; (ii) they are felt to be very significant with respect to the Italian language and/or to our source data. To give an example of (ii), we can quote the case of the Italian verb synset {*rendere, fare, far\_diventare, ridurre*}, meaning "to make someone/something become", which has no corresponding synset in the common selection, but has been kept in our list of bases, since it refers to a really central concept in Italian, as is demonstrated both by the very high number of relations displayed by the verbs which make up it (e.g., the synset has 267 hyponyms) and by the level they have in the taxonomy (they are all tops).

#### 5.4 PSA results for the first subset

The first step of our work on the BCs was to move from single word-senses and create synsets. This was done by firstly associating automatically the following data:

- monolingual LDB synonyms
- monolingual LDB definitions tagged as synonymal definitions
- appropriate semantic indicator synonym data from the bilingual dictionary
- data from our electronic synonym dictionary.

The data obtained were then evaluated and manually tested using the criteria already defined.

The BCs were then grouped together under the preliminary Top Concept headings and these were verified for adequacy.

For each item contained in the base concept synsets, then, we began to semi-automatically collect the first level of hyponym. This has also allowed us to evaluate the synsets built for consistency.

Then we began to assign hyperonyms to BCs, when these are not tops. We use the automatically assigned monolingual LDB hyperonyms as our starting point in this case but it is frequently necessary to create our own hypothesis of hyperonym. For example, our ‘*person*’ taxonomy is extremely shallow: all types of people appear at the first level, from very general classifications (e.g. *woman*), to particular types (e.g. *crybaby*, *slanderer*), to very specific professions (e.g. *glassblower*). We are now attempting to structure the ‘*human occupation*’ taxonomy and we find that we are obliged to intervene on our source data imposing a restructuring by raising certain concepts (such as *worker*, *shopkeeper*, *artist*), locating them at an intermediate level and then linking more specific types of human occupations to these general categories by a hyperonym/hyponym relation. In this case, we are imposing a new structure on existing data. However, to achieve a coherent structure, we will probably also need to introduce a further level in the hierarchy, using a term to represent the idea of human occupation; the problem is that at this level we will probably have to introduce an artificial term ‘*occupazione\_umana*’ which cannot be found in our source data. A similar intervention may also be performed on the ‘*instrument*’ taxonomy: we currently have all kinds of instruments indiscriminately associated at the first level under our base concept, e.g. *musical instruments*, *manual instruments*, *mechanical tools*, *measuring instruments*. We could introduce some artificial multi-words in order to cluster these different groups of tools/instruments. An alternative solution could be to introduce a *role\_instrument* relation, using as values terms such as ‘*measuring*’, ‘*music*’, etc., extracted from our data. It is interesting to note that although our source data contain no multi-words (true multiwords are relatively rare in the Italian language), at the highest levels of our taxonomies it appears that the obvious ‘gaps’ we find could generally be filled by this type of terms.

Finally, all the kinds of other internal relations are also being encoded (when data are easily extractable and/or we feel that a relation is particularly prominent for the description of a word meaning), both by semi-automatically treating information found in our LDB and by intervening manually when necessary. Thus, for instance, particular patterns occurring in the definitions of words in a same taxonomy can often be used to semi-automatically identify certain relations, or particular (recurrent) constructions, implicitly referring to certain relations, can be manually processed to encode the relations themselves.

In the following, more detailed information on the work carried out so far is given both for the verbs and for the nouns.

#### 5.4.1 Results obtained for verbs

As far as the verbs are concerned, we have built synsets for && 153 Italian BCs (131 common to all the sites + 22 drop-outs recovered so far) and stated the equivalence relations to WN 1.5. The synsets built contain about 300 lexical items altogether (but these results are still preliminary).

Up to now, the data related to 48 synsets have been analysed in details: i.e., the sense disambiguation has been performed (cf. above) and data on hyperonymy and hyponymy have been collected. For these synsets, besides encoding the fundamental relations, we have also tried, when possible (i.e., data where present in definitions and/or rather easily identifiable/extractable) to encode information related to other links. In particular:

i) we have extracted cause relations for the hyponyms of BC synsets indicating  $\pm$ causative change-of-state: e.g., all the adjectives occurring as complements of the verbs *rendere* - to make someone/something become - and *diventare* - to become - within definitions refer to states caused by (the events indicated by) the verbs defined and therefore a cause relations was stated between the verb defined and the adjective occurring in its definition:

*ingrassare* = far\_*diventare* *grasso*  
 (to make someone become fat)  
*ingrassare* CAUSES *grasso*

*ingrassarsi* = *diventar* *grasso*  
 (to become fat)  
*ingrassarsi* CAUSES *grasso*

ii) for various taxonomies a recurrent pattern was then used to identify involved\_agents/patients (*detto di* = said of);

iii) other different relations were encoded (involved, involved\_instrument, involved\_source\_direction, involved\_target\_direction, has\_subevent), by using recurrent



patterns displaying regular semantic references within a same taxonomy: e.g., a PP headed by the preposition *da* may have various semantic references; however, within the definitions of the hyponyms of the Italian synset corresponding to {*remove*} it generally refers to the involved\_source\_direction.

Among the BCs analysed, various BCs falling within the 12 subsets identified as the common sets on which to try the automatical vertical ‘extensions’ had been chosen. However, since the whole work on verbs has required more time, due to the recalled disambiguation problems, we are now beginning to perform the automatical extension and we have no data available yet for discussion.

The table below shows the type of relations coded until now for verbs and the relative quantitative data.

*Relations Encoded for the PSA Verbs First Subset*

*Table 25: Relations Encoded for the First Subset Verbs (PSA).*

<b><i>Relations</i></b>	<b><i>Total</i></b>
<b>HAS_HYPERONYM</b>	2276
<b>HAS_HYPONYM</b>	33
<b>INVOLVED</b>	339
<b>INVOLVED_AGENT</b>	3
<b>INVOLVED_PATIENT</b>	78
<b>INVOLVED_INSTRUMENT</b>	15
<b>INVOLVED_LOCATION</b>	1
<b>INVOLVED_SOURCE_DIRECTION</b>	40
<b>INVOLVED_TARGET_DIRECTION</b>	30
<b>CAUSES</b>	622
<b>IS_CAUSED_BY</b>	5
<b>HAS_SUBEVENT</b>	32
<b>IS_SUBEVENT_OF</b>	1
<b>Equivalence to ILI (WN1.5)</b>	114

### 5.4.2 Results obtained for nouns

As far as the nouns are concerned, we have built synsets for the 547 Italian BCs (497 common to all the sites + 50 drop-outs recovered so far) and stated the equivalence relations to WN 1.5.

Then we have worked on the verification and analysis of the first hyponymic level for part of our synsets. During this stage we have encountered various types of problems which had not been foreseen at the moment these internal semantic relations were defined. This has meant that we have been obliged to adopt arbitrary criteria and operational modes which permit us to achieve a reasonable level of coherence and homogeneity; however, as the criteria now being adopted had not having been defined and discussed *a priori* they are liable to alterations and corrections as work proceeds and we are able to verify their usefulness.

For example, we can cite the case of the term *edificio* (building) and its hyponyms. This term automatically selects 64 hyponyms, 8 Part\_of relations and 20 Set-of relations in our main source - a total of 82 items. Analysing the 64 hyponyms and inserting manually those semantic relations which could not be automatically extracted from our data (42 Role\_location, 16 Has\_mero\_part, 2 Has\_holonym) we found that we were forced to establish a set of coherent criteria with respect to the content of these fields. The most frequent of these relations, the role for which the building is destined, is in fact expressed in the definitions by noun, verb and also adjectival phrases.

For example, *abitazione* is defined as follows:

“edificio o parte di esso, *in cui si abita*” (a building or part of a building, in which you live). In this case, the role relation is expressed as a verbal phrase and the content of the Role\_location field should be the verb *abitare*.

However, in the same set of hyponyms, we find *casa* defined as “edificio suddiviso in piani e vani, adibito ad abitazioni”. In this case, the relation is expressed by a noun phrase and thus the content of the Role\_location field should be the noun *abitazione*. In this case, *abitazione* is also a near\_synonym of *casa*.

The inconsistency which emerges is that two near synonyms such as *casa* and *abitazione*, which have the same function, are given two different relations: one to a verb, the other to a noun. Intuitively, it appears preferable to use the verb in both cases, but as these relations are extracted from the definitions given in our source LDB, in this first encoding stage we have decided to keep the data supplied by the definition, i.e. the verb *abitare* for *abitazione* and the noun *abitazione* for *casa*.

Another instance of the lack of coherence in our source definitions, which means that we have to adopt what may appear as an arbitrary criterion, is given by another example

consisting of four hyponyms of *building*: *chiesa*, *moschea*, *sinagoga* and *tempio*. The Role\_location semantic relation which is suggested by the related definitions is the same: they are all *places of worship*. But, whereas for *chiesa* and *tempio* the relation is expressed with the same term *culto* (worship) for *sinagoga* the term used is *rito* (rite) and for *moschea* we have the following definition “edificio sacro dei musulmani” where it is the adjective, ‘sacred’, that suggests the role of the building for purposes of worship. In this case, we have decided to enter the word *culto* as value of the Role\_location relation in all four cases.

There is a clear contradiction implicit in the above two operational decisions: while in the case of *casa* and *abitazione* we have taken the data found in the definition as the value of our semantic relation, in the second case we felt that *culto* was the most representative term and we have thus ignored the definition data in two of the four entries. In both cases, a careful evaluation and manual interventions on the data were necessary.

In most cases, we have used the term given in the definition as the value for our semantic relation, but as is clear from the examples cited above, this can cause problems of inconsistency; it can also mean that at times we are obliged to ignore the semantic relation implicit in the definition as the term used is an adjective - a grammatical category not allowed as a possible value for these relations.

Another problem that has emerged during this phase, is the heavy costs in terms of time involved in inserting these relations in our EWN entries, even when no decision taking is involved. To give a simple example, we can cite the Has\_mero\_part relation. The definitions in this subset generally evidence the role of the building but in some cases many structural elements are also listed and, if we also want to encode all these relations, the work increases considerably. Let us give just one example of this: *castello* (castle) is defined as: “grande edificio munito di mura e torri, circondato da un fossato, in cui abitavano i signori feudali”. We can thus derive:

Role\_location: *abitare* (to live in)  
 Has\_mero\_part: *mura* (walls)  
 Has\_mero\_part: *torre* (tower)  
 Has\_mero\_part: *fossato* (moat)

This means that we have inserted manually four internal relations in addition to the hyponym relation which had been obtained automatically. And *castello* is certainly not an isolated case. Obviously, a scrupulous encoding of this type for all the hyponyms extracted starting from our set of ~700 base concepts is not possible at this stage.

Another serious problem evidenced by the analysis of the hyponyms is that of the lack of depth in our extracted taxonomies which, however, covers an enormous heterogeneity in the elements that have thus been grouped together at the same level by the very generic defining formulae. The most obvious case of this is provided by the set of Agents which in our source database are mainly defined with the formula *chi* (pronoun) + VP. (‘chi = who’

has been mapped to the base concept synset {*persona, individuo, essere\_umano, uomo*}, therefore hyponyms of *chi* are considered as hyponyms of *person*). This means that 2000 items have been automatically tagged as agents. However, if we verify this set of data we find that if it is to be rendered significant a considerable manual intervention is needed in order to distinguish the various types of highly specific agents: *operaio, artigiano, artista, venditore, fabbricante, negoziante* (types of human occupation) - from far more very generic agents such as *comunicatore, ammiratore, calunniatore*, etc. (actions that one can perform) and from also Patients of the type “*chi è affetto da*”, “*chi soffre di*”, etc. (who is affected by, etc.). As can be seen, all this type of redistribution and attribution of not automatically coded relations implies considerable work and numerous decisions which are not always easily taken.

Here below, we give first a few preliminary figures which refer to the twelve subsets that we have agreed to study in depth (see section 2 of this report). We list the main types of semantic relations that are being extracted for each of the ‘extensions’ with preliminary data on the number of entries treated so far for these extensions. Then, a table is given, showing the results obtained in general for the nouns.

### ***1st order entities:***

#### *Human occupations*<sup>5</sup>

No. of base concepts treated: 8 Synsets - 17 items

No. of first level hyponyms processed so far: 154

Main types of semantic relations: hyponym, synonym, set\_of, role\_agent, role\_patient

#### *Places*

No. of base concepts treated: 17 Synsets - 49 items

No. of first level hyponyms processed so far: 750

Main types of semantic relations: hyponym, synonym, part\_of, role\_location

#### *Instruments*

No. of base concepts treated: 7 Synsets - 19 items

No. of first level hyponyms processed so far: 937

Main types of semantic relations: hyponym, synonym, part\_of, set\_of

---

<sup>5</sup>This is a difficult subset for us to extract; our entries relevant to this semantic area are almost all found as hyponyms 'chi' and therefore under the base concept synset that can be identified in 'person'. For this reason, we have to scan the definitions where 'chi' appears in the genus position manually (nearly 2000), recognizing those entries which represent some kind of human occupation. We are also forced to impose some minimum structure by introducing some intermediate hyperonyms (see the discussion above). Here we give some preliminary data.

*Food & drinks*

No. of base concepts treated: 15 Synsets - 26 items

No. of first level hyponyms processed so far: 251

Main types of semantic relations: hyponym, synonym, made\_of,

*Buildings*

No. of base concepts treated: 7 Synsets - 19 items

No. of first level hyponyms processed so far: ~400

Main types of semantic relations: hyponym: 300

synonym: 30

set\_off: 27

part\_of: 16

plus also: role\_location, has\_mero\_part, has\_mero\_madeof, has\_holonym

**2nd order entities:***Physical states or properties (shape, weight, colour, size, taste, etc.)*

No. of base concepts treated: 6 synsets - 13 items

No. of first level hyponyms processed so far: 350

Main types of semantic relations: hyponym, synonym, part\_of, set\_of

*Communication*

No. of base concepts treated: 5 synsets - 16 items

No. of first level hyponyms processed so far: 256

Main types of semantic relations: hyponym, synonym, part\_of, set\_of

*Perception*

Not yet treated

*Feelings*

No. of base concepts treated: 6 synsets - 17 items

No. of hyponyms processed so far: 166

Main types of semantic relations: hyponym, synonym, NtoV

*Movements*

No. of base concepts treated: 8 synsets - 17 items

No. of first level hyponyms processed so far: 228

Main types of semantic relations: hyponym, synonym, NtoV

*Know/knowledge*

No. of base concepts treated: 8 synsets - 21 items

No. of hyponyms processed so far: 460

Main types of semantic relations: hyponym, synonym, NtoV

Table 26: Relations Encoded for the First Subset Nouns (PSA).

<b>Relations</b>	<b>Total</b>
<b>HAS_HYPERONYM</b>	5000
<b>HAS_HYPONYM</b>	
<b>HAS_HOLONYM</b>	6
<b>HAS_HOLO_MEMBER</b>	60
<b>HAS_HOLO_PART</b>	83
<b>HAS_HOLO_PORTION</b>	
<b>HAS_MERONYM</b>	
<b>HAS_MERO_MADEOF</b>	17
<b>HAS_MERO_MEMBER</b>	
<b>HAS_MERO_PART</b>	30
<b>HAS_MERO_PORTION</b>	
<b>ROLE</b>	
<b>ROLE_AGENT</b>	1855
<b>ROLE_INSTRUMENT</b>	196
<b>ROLE_LOCATION</b>	164
<b>ROLE_PATIENT</b>	20
<b>ROLE_TARGET_DIRECTION</b>	
<b>XPOS_NEAR_SYNONYM</b>	13
<b>Equivalence to ILI(WN1.5)</b>	540

### 5.5 Looking for heuristics for automatic encoding

As consequence of the problems we considered above, one of our main objectives will be to look for heuristics to automatically encode as many semantic relations as possible. For this purpose, we are considering:

⟨ Defining *formulae* which can be used to automatically perform the encoding of certain internal semantic relations. For example:

a ‘*da*-phrase’ in the definitions of the hyponyms of *togliere* (to remove) (and its synonyms), always indicates the ‘involved\_source\_direction’;

a ‘*mediante/attraverso*-phrase’ in the definitions in different taxonomies can be connected with the ‘has\_subevent’ relation.

⟨ Adjectives accompanying / qualifying the genus term which could also be useful to identify particular meanings and/or semantic relations:

e.g., all adjectives accompanying either *rendere* (to make someone/something to become) or *diventare* (to become) indicate a state being ‘caused’ (cf. above).

Unfortunately not all homogeneous subsets show the same degree of consistency in their defining *formulae*, as demonstrated by a survey on the two semantic fields of *Instruments* and *Buildings*.

For *instruments* the following *formulae* can be used:

- <i>usato per</i> (used for)	+ VP	68
- <i>atto a</i> (able to)	+ VP	58
- <i>con cui</i> (by means of which) + 3rd sing. Pres.Indic.	14	
- <i>usato in</i> (employed in)	+ NP	7

In this way, the Role\_Instrument can be automatically assigned, taking as value the Infinitive or the Noun following the *formulae*.

In other cases the value of this relation is carried by the Adjective which qualifies the genus term.

For example, the adjective *musicale* allows the coding of 86 Role\_Instrument relations.

On the contrary, although the subset of hyponyms of *Building* does show a few significant patterns, in most cases the definitions differ from each other. While instruments are defined on the basis of their function, buildings are considered, in their definitions, from numerous points of view such as:

- form or structure
- function
- materials they are made of
- type of inhabitants (public buildings)
- types of keepers (shops)
- types of objects which are sold or made or kept in them

This makes an automatic coding of the different semantic relations they are concerned with far more complex.

With respect to the agent/patient terms, we found that a few automatically manageable defining *formulae* could allow the extraction of the related roles, and consequently a better distribution of items within the {*chi*} taxonomy (cf. above) as we show in the table below.

Table 27: Defining Formulae for Agents/Patients (PSA)

<b>patterns</b>	<b>occurrences</b>	<b>agent/patient</b>
<i>chi vende</i>	86	<i>venditore</i> (seller)
<i>chi fabbrica</i>	38	<i>fabbricante</i> (producer)
<i>chi fabbrica e/o vende</i>	18	<i>fabbricante e venditore</i> (seller, producer)
<i>chi lavora</i>	43	<i>lavoratore</i> (worker)
<i>chi esercita</i>	40	(various)
<i>chi fa</i>	161	(various)
<i>chi esegue</i>	15	(various)
<i>chi per professione</i>	7	(various)
<i>chi per mestiere</i>	9	(various)
<i>chi professa</i>	8	<i>promotore</i> (promoter)
<i>chi sostiene</i>	12	<i>sostenitore</i> (assertor)
<i>chi aderisce</i>	2	<i>seguace</i> (follower)
<i>chi appartiene</i>	25	
<i>chi studia</i>	21	<i>studioso</i> (scholar)
<i>chi è affetto da</i>	6	(various)
<i>chi soffre di</i>	5	(various)
<i>chi è malato di</i>	2	(various)

### 5.6 Prospects for the extendibility

In this section of the deliverable, we have described the recent work at Pisa which has involved devising strategies and operational modes in order to begin to encode large quantities of entries as input to our concept net. We have focused our attention on the problems we have to tackle when trying to model our source data to meet the requisites of the project (and in particular to define the most efficient procedures, both manual and semi-automatic, to encode as many semantic relations as possible in our entries). In fact, when treating this core part of our data (the most central concepts), we are obliged to intervene heavily manually in order to ensure coherency in both the structure and the contents. For this reason, we have had to revise our original estimate of the amount of time needed to process all the data regarding the base concepts set and the first level of hyponyms (this, in fact, represents more than 50% of our main resource data). This stage should be completed by the end of June 1997. On the other hand, from the experience acquired during this stage, it appears that the acquisition of entries which are lower down or at the bottom of the taxonomies will be a far more rapid procedure - these tend to be unambiguous terms and can normally be selected automatically by their hyperonyms, without the necessity for a heavy manual intervention.



## 6. First Building Phase. FUE

### 6.1 Tools and Resources

#### 6.1.1 Databases as Input

MRDs and Databases we use or have available to use in the next phases of the project are the following:

- Spanish Monolingual:
  - DGILE- Diccionario General Ilustrado de la Lengua Española (not used in this 1st. phase)
- English/Spanish Bilinguals:
  - VOX-HARRAP'S Esencial Eng-Span
  - VOX-HARRAP'S Esencial Span-Eng
  - VOX-HARRAP'S Advanced Eng-Span (used but not generating results in this 1st. phase yet)
  - VOX-HARRAP'S Advanced Span-Eng (used but not generating results in this 1st. phase yet)
  - Mergings of Bilinguals detailed above (not used yet)
- PIRAPIDES Verbal Database
  - Database developed within the PIRAPIDES Project (University of Barcelona). Consists of 3600 English verb forms organized around Beth Levin's Semantic Classes connected to WN1.5 senses. Theta-Grids specifications for each verb. Ambiguous translation to Spanish forms (under process of disambiguation). Diathesis information for each verb under development.
- WordNet 1.5

The following table contains the figures of both nouns and verbs (at entry and sense level) that occur within the Vox DGILE monolingual dictionary. These figures can be considered as an upper bound for our approach, taking into account that most of the other sources have a more limited coverage.

Table 28: Number of entries and senses (FUE)

	<i>nouns</i>	<i>verbs</i>
<b>entries</b>	65000	11000
<b>senses</b>	105000	24000

Next we present in the following table the content of taxonomic relations that are explicitly defined in our lexical sources, i.e. those relations that can be obtained directly from the source. In this case the figures do not imply an upper bound because other relations could be extracted by means of more complicated procedures.

Table 29: Taxonomic data per sense (FUE)

Type of relations	Noun Senses	Verb Senses
<b>ANTONYM</b>	260	20
<b>BE_IN_STATE</b>		169
<b>CAUSED_BY</b>		14
<b>CAUSES</b>		50
<b>HOLONYM</b>	2466	
<b>HYPERONYM</b>	111624	
<b>INVOLVED</b>		695
<b>MERONIMY</b>	3756	
<b>SYNONYM</b>	18736	1400
<b>XPOS_NEAR_SYN</b>	7800	

### 6.1.2 Local Tools and Technology

The following tools have been developed within the Project:

- A System for automatically mapping Spanish words to WN1.5 Synsets via Bilingual MRDs
- VI (Verification Interface): A tool for revision and validation of SpanishWord-WN1.5 Synsets automatic mappings
- Automatic Taxonomy Builder from monolingual MRDs (not generating results yet in this 1st. Phase)

General Data are managed using UNIX platforms

## 6.2 Local Approach

FUE approach follows the building guidelines of EWN but differs from AMS and PSA methodology in one issue. The basic strategy to build the Spanish WordNet is mapping Spanish words into WN1.5 synsets using automatic and semi-automatic methods. Thus, instead of building language-specific structures from scratch and then matching them to ILI (which is a variant of WN1.5), FUE first resembles WN1.5 structure and then adapts this structure to Spanish-specific conceptualizations.

The assumption behind this methodology is that, being wordnets something close to conceptual databases, English and Spanish are close enough to have globally compatible conceptualizations. There are also practical reasons for such approach: FUE's available resources for Spanish are not rich enough to build a wordnet-like structure from scratch.

Advantages and drawbacks for FUE approach are evaluated below in Section 3.5.

Summing up, FUE's general approach is as follows:

The Spanish WordNet will largely map, at least in the first stages of development, the organization of WN1.5. This involves mainly working on the following semantic relations: hypernymy-hyponymy, synonymy, antonymy (verbs and nouns), part-whole relations (nouns) and causation (verbs). Spanish words are assigned to WN1.5 synsets using automatic and semi-automatic methods. Nevertheless, the central part of 1st. building phase, namely Base Concepts and surroundings, is either manually built or manually revised.

In further phases this WN1.5-like organization will be completed and/or modified by:

- Merging the WN1.5-like network to taxonomies coming from the Monolingual MRD
- Acquiring and incorporating to the network semantic relations defined within EWN which can not be imported from WN1.5

Work in this first phase has proceeded as detailed below.

### 6.2.1 Nouns

Automatic Methods- Background Tasks (this work is described in D005):

- Eight methods for automatic Spanish-to-WN1.5 mapping have been developed and manually tested using VI. Four methods have proved to be reliable (accuracy from 85.8% to 93.2% estimated by manual inspection). The only source used at this stage is VOX-HARRAP'S Esencial
- More methods and exploitation of the rest of MRDs available are being developed at the moment
- DGILE Monolingual has been fully converted into a lexical taxonomy automatically. The results have been evaluated being the conclusion that this methodology is not useful to fully build a semantic network but it is to extract partial taxonomies for the lower nodes; e.g. the taxonomy for 'food' is not reliable but that for 'fruit' it is.
- Some efforts have been devoted in this phase to prepare some other sources (DGILE, VOX-HARRAP'S Advanced) to be used in further phases.

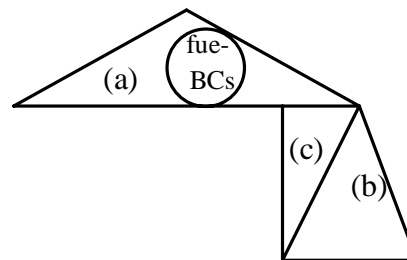
#### ***Manual work in this 1st. Phase: Base Concepts and Surroundings, namely the 'Core' of the Spanish WordNet.***

Work in this phase has focused in building a reliable core of the Spanish WordNet from which to extend in the next phases. The Core maps for Spanish the upper nodes of WN1.5. It bears the following characteristics:

- Synsets are complete, i.e. all synonymy relations are present
- Synsets are fully linked by hypernymy up to the top, so there are neither 'floating' nodes nor missing intermediate nodes between any synset and the top -being the top the replica of WN1.5 'noun.top' file: 30 synsets hierarchically organized
- Relations other than synonymy and hypernymy-hyponymy have as well been imported from WN1.5 and tested at this stage
- The Core includes all Base Concepts but as well any hypernym of those up to the top.

The Core consists of the following subsets (see schema below):

- WN 1.5 noun.top file synsets
- Any hyponym (one level) of those
- Base Concepts missing in the local approach (i.e., coming from intersections of the other sites which not intersect with original Spanish BCs) except for those already present in the two subsets detailed above
- Hypernyms of these 'missing' BCs up to the top



The **NCORE** consists of:

- (a) TOPS - all WN1.5 noun.tops file plus one level of hyponyms of them total 628 synsets). All FUE-BCs (361 synsets) are included here.
- (b) BC\_INTERSECTIONS - synsets coming from intersections of the other sites. These concepts were not present in FUE-BCs.
- (c) HYPERNYMS - since (a) is the full top-block of the hierarchy and (b) are concepts scattered somewhere below, the full chain of hypernymy from (b) to (a) has been built.

All synsets in the Core have been built manually by a lexicographer by: (i) translating WN1.5 synset variants to Spanish; and (ii) extending the resulting Spanish synsets with their synonyms. Nevertheless testing of automatic methods already provided varint candidates for a number of synsets. These have been furtherly manually revised and then incorporated to the corresponding synset

### ***Automatic Work in the 1st. Phase: First Extension***

Applying the four reliable automatic methods described in D005 to the first MRD source gave as output a set of SpanishWord-Synset pairs not present in the Core. The First Extension consists of this set.

In this case the synsets are not complete

In this case the hypernymy lines up to the top are not complete as well, so there are intermediate nodes missing

In the meantime those missing intermediate nodes are skipped, so synsets are linked by hypernymy to the appropriate lower node present in the Core

In this case the data is not 100% reliable so any word bears an index of reliability ranged between 85.8 and 93.2% depending on the method they come from

## 6.2.2 Verbs

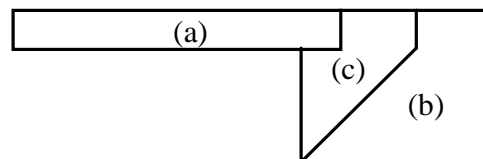
Work in verbs in this phase has been all manual. Most work on verbs during the project will be mainly manual except from importing relationships from WN1.5 and other sources. This is due to (1) the availability of PIRAPIDES database that seems to be more reliable than data coming from bilingual MRDs; and (2) the dubious utility of verbal taxonomies derived from monolinguals.

As in the case of nouns, the database for verbs also maps the WN1.5 organization. In this case, as it is known, the taxonomy is somewhat ‘flat’: the hypernymy lines lead to about 700 top concepts which in WN1.5 are organized in 15 semantic clusters. The EWN equivalent to this clustering is provided by links to the EWN Top Ontology.

### *Work on Base Concepts and Surroundings: the ‘Core’ for verbs*

The Core consists of the following subsets (see schema below):

Original Spanish Base-Concept synsets  
 Base-Concepts coming from intersections of the other sites not intersecting to the Spanish set  
 Hypernyms of these ‘missing’ BCs up to the top



The **VCORE** consists of:

- (a) FUE-BaseConcepts -Original FUE-BCs. All of them belong to the top level of the WN1.5 verb hierarchy.
- (b) BC-Intersections -Synsets coming from intersections of the other sites. Some are top-level in WN1.5, and some are not
- (c) HYPERNYMS -Chain of hypernymy from those synsets in (b) which are not top-level in WN1.5 to the top

The verbal Core bears the same characteristics than that of nouns described above, i.e. full synset, complete hypernymy chain, no ‘floating’ nodes (except for isolated tops), no missing taxonomically intermediate nodes. All synsets in the Core have been built up manually by a lexicographer as described for nouns

### ***First Extension***

It comes from manually disambiguating the PIRAPIDES DataBase described above in 3.1.1. The result is a set of Spanish Word- WN1.5 synset pairs.

As for nouns, in this case synsets are not complete

As for nouns, missing intermediate nodes are skipped in the meantime

Accuracy of the resulting data is still under evaluation

### ***6.2.3 Work on extending and refining the data***

#### ***Other Relations***

In this phase relations other than synonymy and hypernymy-hyponymy have been imported from WN1.5. Results have been evaluated as it will be detailed below. As a result of the evaluation some corrections have been made, basically:

- Nouns: some incorrect mappings have been removed
- Verbs: the WN1.5 relation ‘entailment’ has been redefined manually to give EWN causation relations

Moreover to show and prospect the capabilities of the framework defined within EWN a limited semantic subset involving ‘communication’ concepts has been manually worked out. This semantic area is specially confusing and badly solved in WN1.5. Concepts have been linked by using several relations other than those imported from WN1.5. This work will be detailed below.

### ***6.2.4 Top-Ontology Clustering***

Verbal and Nominal Cores have been (manually) linked to the EWN Top-Ontology in the following way:

- Any top node in the Spanish taxonomies have a link to the interlingual Top-Ontology
- Nodes below those have a link to the Top-Ontology in the case they could not inherit his parent’s link so they need to be linked to a more specific (lower) Top-Ontology concept

### 6.3 FUE BC Selection

This work has already been described in D005 and D006 deliverables. Criteria for selecting a word as a Base Concept for Spanish have been the following:

1. Word is a translation of either a top concept or a direct hyponym of a top concept in WN1.5, and moreover either (2) or (3):
2. It occurs as genus word in the DGILE monolingual MRD 5 or more times
3. It shows a high frequency of occurrences in corpora: either (3.1) or (3.2):
  - 3.1 It occurs 50 times or more in the DGILE definitions corpus.
  - 3.2 It occurs 100 times or more in LEXESP (a 3.000.000 word unrestricted corpus)

The result of the selection was 361 noun synsets and 104 verb synsets. All of them still are within the Core, so FUE has no ‘drop-outs’. The only thing which has been done with respect to the final set of Base Concepts has been to incorporate ‘missing’ BCs -those coming from intersections of the other sites.

### 6.4 FUE Results

#### 6.4.1 Figures

We present in the following tables the result of the first subset differentiated for Core and Core + 1st extension.

Table 30: Noun First Subset differentiated for Core and Core + 1<sup>st</sup>. Extension (FUE)

Core	Total 1st. Subset (Core + 1st. Extension)
810 synsets, corresponding to:	7983 synsets, corresponding to:
1071 number of entries (words)	7896 number of entries (words)
1271 number of senses (variants)	9722 number of senses (variants)

Table 31: Verb First Subset differentiated for Core and Core + 1<sup>st</sup>. Extension (FUE)

Core	Total 1st. Subset (Core + 1st. Extension)
222 synsets, corresponding to:	1767 synsets, corresponding to:
279 number of entries (words)	1269 number of entries (words)
416 number of senses (variants)	2813 number of senses (variants)

Table 32: First Subset differentiated for Core and Core + 1<sup>st</sup>. Extension (FUE)



<b>Core</b>		<b>Total 1st. Subset (Core + 1st. Extension)</b>
1032	synsets, corresponding to:	9750
1350	entries (words)	9165
1687	senses (variants)	12535
		synsets, corresponding to:
		entries (words)
		senses (variants)

Next we present the figures corresponding to relations extracted, including the internal synonymy and the top concept links.

Table 32: Relations Encoded for Core and Extension (FUE)

Language Internal	Nouns	Verbs	Nouns	Verbs
<b>synonymy:</b>				
X variants per synset	1.56	1.87	1.21	1.59
<b>ANTONYM</b>	32	18	127	129
<b>BE_IN_STATE</b>				
<b>CAUSES</b>		9		60
<b>HAS_HYPERONYM</b>	808	88	8158	1436
<b>HAS_HYPONYM</b>	808	88	8158	1436
<b>HAS_HOLONYM</b>				
<b>HAS_HOLO_MADEOF</b>	2		9	
<b>HAS_HOLO_MEMBER</b>	4		33	
<b>HAS_HOLO_PART</b>	23		310	
<b>HAS_HOLO_PORTION</b>				
<b>HAS_MERONYM</b>				
<b>HAS_MERO_MADEOF</b>	2		9	
<b>HAS_MERO_MEMBER</b>	4		33	
<b>HAS_MERO_PART</b>	23		310	
<b>HAS_MERO_PORTION</b>				
<b>HAS-SUBEVENT</b>		1		1
<b>INVOLVED</b>	1	5	1	5
<b>INVOLVED_PATIENT</b>				
<b>IS_CAUSED_BY</b>		9		9
<b>IS_SUBEVENT_OF</b>	1		1	
<b>NEAR_ANTONYM</b>				
<b>NEAR_SYNONYM</b>	2	6	2	6
<b>ROLE</b>	6		6	
<b>ROLE_AGENT</b>				
<b>ROLE_INSTRUMENT</b>				
<b>ROLE_LOCATION</b>				
<b>ROLE_PATIENT</b>				
<b>XPOS_NEAR_SYNONYM</b>	1	1	1	1
<b>Equivalence to ILI (WN1.5) eq_synonym:</b>	810	222	7983	1767
<b>Top Concept Links</b>	444	194		

Next we present the figures corresponding to the top concept links classified by cluster. Under the title "direct ass" we present the links explicitly defined, for nouns and verbs, while in the other two columns the figures corresponding to the inherited links are provided.

Table 32: Hierarchical overview of distribution of sense per (main) cluster (FUE)

TopConcept Cluster	direct		indirect		
	N	V	N	V	
[Lit] Top-0	1		2		
Hyponym					
[Lit] FirstOrderEntity-0	2		11		
Hyponym					
[Lit] Functional-0	33		444		
[Lit] Group-0	28		256		
[Lit] Object-0					
Hyponym					
[Lit] AnimateObject-0	3		21		
Hyponym					
[Lit] Animal-0	12		233		
[Lit] Creature-0	2		12		
[Lit] Human-0	27		863		
Hyponym					
[Lit] HumanKinship-0	2		53		
[Lit] HumanNation-0	1		3		
[Lit] HumanOccupation-0	14		235		
[Lit] Plant-0	18		221		
[Lit] InanimateObject-0	16		294		
Hyponym					
[Lit] Building-0	6		113		
[Lit] Container-0	4		100		
[Lit] Furniture-0	3		33		
[Lit] Garment-0	3		84		
[Lit] Instrument-0	12		373		
[Lit] Vehicle-0	7		94		
[Lit] Weapon-0	1		24		
[Lit] Part-0	15		181		
Hyponym					
[Lit] BodyPart-0	11		108		
[Lit] Place-0	5		116		
[Lit] Portion-0	4		48		
[Lit] Substance-0	35		308		
Hyponym					
[Lit] Gas-0	1		10		
[Lit] Liquid-0	5		55		
[Lit] Solid-0	13		105		
[Lit] Symbol-0	2		25		
Hyponym					
[Lit] ImageSymbol-0	4		24		
[Lit] LanguageSymbol-0	20		231		
[Lit] MoneySymbol-0	4		21		
[Lit] Type-0	2		22		
		N	V	N	V
[Lit] HighOrderEntity-0			1		5
Hyponym					
[Lit] Dynamic-0	13	215	11		559
Hyponym					
[Lit] Activity-0	3	1	1		201

	Hyponym					
	[Lit] ArtActivity-0	7		48		
	[Lit] FightActivity-0	1	12	3	18	
	[Lit] GameActivity-0	1	1	3	44	
	[Lit] LoveActivity-0		1		10	
	[Lit] RecreationActivity-0					
	[Lit] ScienceActivity-0		5		63	
	[Lit] WorkActivity-0		5		65	
	[Lit] BehavioralEvent-0		3		5	
	[Lit] CareEvent-0	2	6	3	23	
	[Lit] Change-0			11	109	
	Hyponym					
	[Lit] CausalChange-0		5		18	
	[Lit] ExistentialChange-0	2	62	2	22	
	[Lit] PhysicalChange-0	9	56	1	6	
	[Lit] PossessionChange-0	16	86	1	3	
	[Lit] QualityChange-0		3		21	
	[Lit] QuantityChange-0	3	47	3	22	
	[Lit] CommunicativeEvent-0	14	104	6	109	
	[Lit] ConsumptionEvent-0		1		23	
	[Lit] EducationEvent-0					
	[Lit] ManagementEvent-0	4	13	1	5	
	[Lit] MentalEvent-0	19	46	4	71	
	[Lit] MotionEvent-0	23	369	10	100	
	[Lit] OperationalEvent-0	3	5	1	181	
	[Lit] SoundEvent-0	2	26	4	53	
	[Lit] Manner-0		9		42	
	[Lit] Perception-0	8	38	3	79	
	[Lit] Phenomenon-0		4		59	
	Hyponym					
	[Lit] WeatherPhenomenon-0	2		18		
	[Lit] Stative-0	13	36	6	199	
	Hyponym					
	[Lit] ExistentialState-0		1	7	5	137
	[Lit] LocationState-0	3	21	5	14	
	[Lit] Meaning-0	2	4	5	200	
	[Lit] MentalState-0	9	27	22	334	
	[Lit] ModalState-0	7	4	5	44	
	[Lit] PhysicalState-0					
	Hyponym					
	[Lit] PhysicalColour-0		1		21	
	[Lit] PhysicalQuantity-0		4		83	
	[Lit] PhysicalShape-0		5		88	
	[Lit] PhysicalSize-0					
	[Lit] PhysicalSmell-0					
	[Lit] PhysicalTemperature-0					
	[Lit] PhysicalTexture-0					
	[Lit] PhysicalWeight-0					
	[Lit] QualityState-0			11	201	
	Hyponym					
	[Lit] CausalRelation-0			4	15	
	[Lit] PossessionRelation-0	2	9	2	54	
	[Lit] SocialState-0			7	56	
	[Lit] Time-0	6	10	6	113	

### **6.5 Amount of resources spent**

855 hours/person or 5.3 person/months

Only work on producing and tuning data has been considered. Tasks on building software and preparing, developing and evaluating experiments is not taken into account.

### **6.6 Quality of the data. Evaluation.**

The Spanish Database as it is now has been evaluated in order to determine the quality of the resulting data. This evaluation has only been done with the 'Cores' so data coming for the First Extension are not considered throughout the rest of the section.

As explained above the Spanish database largely maps WN1.5 structure, consequently any consideration about the quality of this resulting database has to consider two points:

- 1.- adequacy of the WN1.5 structure;
- 2.- quality of mappings from Spanish to WN1.5

- 2.1.- quality of lexical equivalences
- 2.2.- adequacy of imported relations

With respect to (1) this is possibly not the place where to do this kind of evaluation. Different kind of criticisms can be put on WN1.5 being the most relevant too much sense fine-grainedness, artificial multiplication of senses, lack of cross-POS relationships and excessive simplicity of relational information. Undoubtedly, as it is in this first stage, any criticism which apply to WN1.5 also apply to the Spanish WordNet.

Another point to be considered is that the WN1.5 organization is not purely lexical, but lexical-conceptual. That is, synsets stand for concepts which in some cases involves using not only words but also co-locations (multi-word expressions) to express concepts. The Spanish WordNet is as well so. Moreover, differences in lexicalization between English and Spanish entails that some synsets which are realized in WN1.5 by means of words are in Spanish by means of co-locations -and the other way round. E.g. in both WN1.5 and the Spanish WordNet a concept as 'car' will be hyponym of 'motor\_vehicle', not a direct hyponym of 'vehicle'. Which kind of organization (either purely lexical or lexical-conceptual) suits better for IR and other language technologies purposes should be evaluated.

Anyway FUE assumes that the drawbacks of mapping WN1.5 structure have as counterpart some advantages.

First, WN1.5 organization is well established and widely used within the NLP community. Furthermore, most organizative effort comes already done and it is a steady basis to work within along the rest of the project -no organizative model redefinitions will be needed, only improvements and extension of the data. Second, relations used by WN1.5 are few and indeed they don't cover all kinds of linguistic information which can be potentially coded in a lexical database. For instance, some linguistic information that FUE has available for verbs: Levin class assignment, Theta-grids for any word-sense and Diathesis information, can not be encoded within the project. Nevertheless FUE assumes that those relations (hypernymy, synonymy, antonymy, part-whole and causation) are probably the most basic ones to build a relational database around. Third, FUE approach aims to ensure non ambiguous cross-linguistic (Spanish-English) equivalence for all data generated throughout the project -synset by synset. And last, this approach seems to FUE to be the one which, given the time and resources available, will allow to build a semantic database which be (i) large, and (ii) fully cross-linguistically (Sp. to Eng.) linked.

With respect to point (2) the Spanish WordNet structure as it is now has been evaluated resulting in the considerations detailed below.

About the quality of lexical equivalences (2.1.) those of the core are fine, since they have been manually done. (See examples of taxonomies built in the appendixes -only hyponymy is represented there). The quality of automatic (nominal) mappings for extesions are, as explained above and in D005, subject to an index of reliability. Much work is being done in the direction of improving this point so it is expected to reach a level of accuracy of automatic methods at least above 90% for the whole final database.

The main problem which has been detected on the approach with respect to this point is that, due to similarity of meaning, some Spanish words are sometimes mapped to a synset and as well to its hypernym (or hyponym). These cases have to be traced manually. In some cases duplication has been maintained since corresponded to different meanings of the same word. In other cases duplication has been proved to be an error. In the Core 8 cases (among 810 synsets) have been detected (and corrected) for Nouns. 3 cases of this problem have been detected for Verbs. A couple of examples can be seen below, where, in one case 'empleado' has been removed from the upper synset, and in the other 'embarcación' has been replaced by 'nave'.

00002403-n {entidad} ==> 00004473-n {agente\_causal,causa}  
 00004865-n ==> {alma,humano,individuo,mortal,persona,ser\_humano} ==> 05856677-n  
 {**empleado**,trabajador} ==> 06069879-n {empleado}

02009476-n {instrumento} 01991412-n ==> {medio\_de\_transporte,transporte} ==>  
 03233330-n {vehículo} ==> 03235595-n {**embarcación** --> **NAVE**} ==> 02167572-n  
 {embarcación}

With respect to accuracy of relations imported from WN1.5, an evaluation has also been made resulting in that, for Nouns, all imported relations work fine. This is not surprising since relations in WN1.5 stand between concepts (except for antonymy; nevertheless, this relation, for the time being, stands between synsets in the Spanish database) and concepts are by definition non language dependant so they can be non-problematically exported to other conceptual databases.

The main problem is with the WN1.5 relation ‘entailment’ which does not directly correspond to any EWN relation. Subevent and (factive and non-factive) Causation are candidates to map WN1.5 ‘entailmentte’. This has been done manually (see a couple of examples below)

00620218-v {perder} hypernym Activity **entails** --> **is\_caused\_by/nonfactive**  
 00605050-v {competir,contender,medirse}  
 # lose, fail to win => compete, vie, contend #

01203891-v {experimentar,tener} hypernym Perception **entails** --> **is\_caused\_by**  
 01202814-v {percibir,sentir}  
 # experience, receive, have, get, undergo => feel, sense, perceive, pick up  
 #

Summing up, drawbacks and potential benefits of FUE’s approach are the following:

### Drawbacks

- Criticisms to WN1.5 also apply to Spanish-EWN database
  - Too much sense fine-grainedness
  - Lack of cross-POS relationships
  - Simplicity of relational information
    - But: Acquisition of more relationships is not a matter of this phase. Will be worked later.
  - Not pure lexical, but lexical-conceptual database
    - But: Maybe this is not a drawback
- Words double-mapped to contiguous synsets
  - But: Only in 1% of the noun synsets. Manually correctable
- ‘Entailment’ can not be imported from WN1.5 to EWN
  - But: Few cases (4% of the synsets). Manually workable.

## Benefits

- Network organization well established and widely used within the NLP community
- Most networking effort already done
- Steady basis to work within along the rest of the project. Only improvements and extensions required
- Relations present in WN1.5 are the most basic ones
- Ensures full compatibility with the ILI, thus it facilitates multilingual usage
- The approach seems to allow, given limited time and resources available, to build a large and fully cross-linguistically (Spanish to English) linked semantic database

### 6.7 Prospects for the extendibility

Research on automatic methods for building the nominal part of the database is going on. Manual disambiguation of the verbal PIRAPIDES database is so. It is expected that:

- In the next months the Spanish verbal database will be complete and related by the relations importable from WN1.5
- Expectations for the nominal database (about 35.000 synsets cross-linguistically linked) will be reached under a global level of reliability around 90% of accuracy for the extensions (100% for the upper concepts -see below)
- Completion with further relations will be developed in phases to come

Moreover, it has been detected that only 4743 of the total of WN1.5 noun synsets have 5 or more hyponyms. So these may be seen as the upper-level 'big-core' of WN1.5 and the rest (up to 60.000 aprox.) the lower-level 'leaves'. We are committed both to fully cover such 'big-core' and to have it both manually revised and directly mapped from Spanish to WN1.5. At least 1123 of those synsets are already covered now. In order to go on increasing the covering of the 'big-core', any further experiment on methods of automatic mapping to WN1.5 will be made on the set of synsets belonging to the remains of the 'big-core' which still have not been covered at the step. This will entail manual testing on them, hence a stepwise increasing covering of such 'big-core' synsets with a 99% degree of reliability.

## 7. First Building Phase Sheffield

### 7.1 Tools and Resources

The tools that Sheffield has used in this phase are the ConceptNet Toolkit, the Microsoft Access relational database package, and various UNIX-based utilities.

This is a list of resources Sheffield has in electronic format. Not all of these have been converted into a useable format.

- Monolinguals
  - **LDOCE**                    76060 lemmata
    - 21400 nouns
    - 7361 verbs
    - 7333 adjectives
  - **COBUILD**                64908 lemmata
    - 6566 nouns
    - 4962 verbs
    - 3490 adjectives
- Other Data
  - **CELEX**                    52447 lemmata
    - 29494 nouns
    - 8504 verbs
    - 9185 adjectives
  - **COMLEX**                39493 lemmata
    - 21871 nouns
    - 5660 verbs
    - 8170 adjectives



## 7.2 Local Approach

Sheffield has in this first phase concentrated on the semi-automatic extractability of near synonymy relations between nouns and verbs on the basis of morphological relatedness. Derivational patterns consisting of a verbal base form with a suffix, and of a conversion (zero-derivation) of a verbal base form were first taken into account. The matching process consisted of the following stages:

1. extraction of all CELEX lemmata with the appropriate derivational analyses.
2. extraction of the base verb for derivations with suffixes.
3. matching of deverbal nouns with WordNet.
4. matching of the (extracted) base verb with WordNet1.5 and Celex to obtain the WordNet synsets the verb is a member of and the Celex lemma number connected with the verb which gives access to other types of linguistic information contained in the Celex database.

The following types of deverbal nouns were found in CELEX:

---



---

suffixation: (Noun+) Verb + Affix	->	Noun	(2996 in CELEX)
conversion: Verb	->	Noun	(1749 in CELEX)

The following criteria has been used when creating relation links in relation to different types of mapping between nouns and base verbs. *l* and *m* represent '1 sense' and 'more than 1 sense' respectively.

l - l	relation valid by default
m - l	relation valid by default
l - m	disambiguation necessary
m - m	disambiguation necessary

After the mapping process the scope of the relation was extended to other members of the synsets which contain the related noun and base verb.

Apart from cross-part-of-speech relations some work has been done until now on other types of relations within the sets of base concepts.

All top concept relations proposed by AMS have been checked and commented on where necessary.

### 7.3 SHE Base Concepts selection

The criteria that have been taken into account for the selection of the WordNet base concept set are:

- 1 Density
- 2 Hyponymic branching factor
- 3 Position in hierarchy

A measure has been developed which involves all these criteria.

One way to select appropriate WordNet synsets is by means of the density criterion. In our study this relies completely on the notion of hyponymy. The concept of density can be regarded as:

- a) a node's total number of hyponymic descendants;
- b) a node's mean branching factor (mean number of hyponyms in WordNet);
- c) a function of a) and the node's relative position in the hyponymic hierarchy.

Three strategies involving density measures have been considered for nouns and verbs:

The first hypothesis was that synsets of interest display a higher hierarchical (hyponymic) branching factor than the mean branching factor of all noun/verb synsets. The mean branching factor can be computed as the division of the total number of branches in the hyponymic structure of WordNet, grouped by POS, by the total number of synsets belonging to that POS.

For verbs the equation looks like this:

$$\frac{\text{total no. of branches } 10817}{\text{total no. of synsets: } 11363} = 0.95$$

The conclusion from this is that this measure is not usable for verbs because all verb synsets with 1 hyponymic relation or more (2917 in total) can be regarded as candidates for inclusion in the selection set. This criterion would only exclude the 8446 verb synsets that do not have any hyponyms.

For nouns the situation is as follows:

$$\frac{\text{total no. of branches } 47110}{\text{total no. of synsets: } 60557} = 1.829$$

2419 noun synsets have a branching factor of 1, whereas 11028 synsets have a branching factor of two or more. T Also for the nouns the conclusion must be that the criterion is not good enough for selecting a proper subset of noun synsets.

Then a second strategy was followed which examines the local density per node as defined in a) above:

(the total number of local descendants)

-----

(the total number of noun synsets)

Selecting candidates according to this measure has the obvious disadvantage that nodes high up in the hyponymic hierarchy will automatically be more likely candidates, because they dominate more synsets than nodes lower in the structure.

An attempt to balance this out is to take into account the level the concepts occupies within the hyponymic hierarchy the concept is part of, and the length of the different hyponymic chains in WordNet:

total no of descendants

-----

level of concept / total no of levels of the chain  
including the concept

By applying this measure nodes lower in the structure but with a fair amount of descendants will become more likely candidates for selection. Another advantage is that including the total number of levels of the hierarchical chain the concept is part of tends to diminish the over-representation of very detailed hierarchical structures in WordNet (such as the large set of botanical terms).

This measure seemed to be the most promising of the ones under examination. Extracting the 20% topmost values for nouns yielded 1090 distinct noun synsets. For verbs 197 distinct verb synsets came out.

Until now 14 new verbal base concepts have been selected from the Sheffield dropouts following the criterion that any extensions of the original base concept set cannot be hyponyms of synsets which already belong to the sets of base concepts. Using this criterion for nouns did not yield any new base concepts from the dropout set. From the set of Sheffield dropouts candidate verb and noun concepts have been selected manually with respect to their saliency as core concept. These might be added in a later stage to the set of base concepts when other criteria will be used for inclusion, e.g. corpus frequency.

The following table gives a complete overview of the base concept set of Sheffield:

*Table 33: Final Set of Base Concepts (SHE)*

<b>Initially</b>	<b>Nouns</b>	<b>Verbs</b>	<b>Total</b>
<b>Selected</b>	768	300	1068
<b>Missing</b>	231	171	402
<b>Drop-Outs</b>	1232	344	1576
<b>Finally</b>	768	314	1082

#### **7.4 Results**

The results of this mapping process are listed in the following table. Because the manual disambiguation process has not yet been completed yet the figures are an approximation.

*Table 34: Relations Encoded for the First Subset (SHE)*

<b>Language Internal</b>	<b>Nouns</b>	<b>Verbs</b>
<b>XPOS_NEAR_SYNONYM</b>	150 synsets	470 synsets
	350 variants	1550 variants
<b>Equivalence to ILI (WN1.5) eq_synonym:</b>	150	470
<b>Top Concept Links</b>	501	133

Applying the criterium ‘if a dropout noun/verb synset is not a hyponym of any of the base concept synsets it has to be added to the set of base concepts’ yielded 14 new verb synsets and no new noun synsets

Quality of the data if automatically generated:

Experience has shown that the cross-part-of-speech relations generated semi-automatically are for the greater part ambiguous and will need manual processing (see below).

Amount of resources spent: approximately 1 person month.

### ***7.5 Prospects for the extendibility.***

The semi-automatically extracted nominal derivations from verbs by means of suffixation or conversion can automatically be extended to the whole of WordNet1.5 for the monosemous derivations that are related to verbs that are member of only one synset (774 in total). Where a noun is polysemous or the related base verb is a member of more than one synset disambiguation needs to take place. This will be a substantial task. For example 17000 possible relations have been extracted for polysemous conversions that are derived from polysemous verbs.

A complete overview of the records containing noun-verb relations that need disambiguation:

39506 noun synset to verb synset xpos\_near\_synonym relations:

1 - 1	361
1 - n	413
1 - n	4247
n - n	34485

38732 need to be disambiguated manually or semi-automatically

The possibility of automating this disambiguation task is under investigation. One strategy is to take all the members of the synset each deverbal noun is a member of. Then establish whether any other synset member has a derivational relation with another base verb. If there is one, select the synsets from the ambiguous set of synsets the original base verb is a member of which also contain this second base verb. This will at least reduce the number of appropriate synsets for this near-synonymy relation.

Other types of derivation will also be taken into account. One important class are the cross-part-of-speech relations between verbs and nouns on the one hand and adjectives on the other, respectively defined by the EuroWordNet relations STATE of and BE IN STATE. Other information available in CELEX and LDOCE like arity and semantic subcategorisation information for verbs can be used in a later stage to interpret the argument roles of deverbal head nouns within derivational compounds like 'aircraft carrier'. This will yield information for the ROLE and INVOLVED relations.

## 8. Conclusions

In this document we have described the work done for the first subset in EuroWordNet. The work covers a period of 3 months. The work will continue in the next 6 months. We expect that the total first subset and the corresponding ontology of TopConcepts will be available around September 1997. We have focussed the work on a set of common Base Concept which was determined using similar criteria across the different resources. The Base Concepts have the most relations and occupy the most important positions in the hierarchies. This coding of these word meanings is crucial for the quality of the wordnets as a whole. Nevertheless, these words are very polysemous and poorly-defined in traditional resources. Therefore, we have concentrated the manual work on these words and their direct semantic context. This context minimally consists of the synonyms, hyperonyms and hyponyms of the BCs and where possible also other non-hyponymy relations. This document will be updated when the first subset is completed. After September 1997 we will extend this first subset to the full set of 33,000 synsets or 50,000 word meanings in total.

## References

- Boersma, P.  
1996 The Sift Lexical DataBase, Computer Centrum Letteren, University of Amsterdam, June 1996, Sift deliverable D10b, D21a, LRE 62030.
- Carroll, J.  
1990 The Acquilex Lexical Database, Cambridge University.
- Cuypers, Ilse and Pedro. L Díez-Orzas  
1996 Manual of the Novell ConceptNet MS, Internal Report, Novell Begium N.V.
- Díez-Orzas, Pedro. L  
1996 WordNet1.5 Contents and Statistics, Novell LTD. Internal Report, Antwerp.
- Procter, P.  
1978 The Longman Dictionary of Contemporary English, Longman, London.
- Martin, W. and J. Tops  
1986 Het Groot-Woordenboek Nederlands-Engels, Van Dale, Utrecht
- Martin, W. and J. Tops  
1989 Het Groot-Woordenboek Engels-Nederlands, Van Dale, Utrecht
- Vossen P.  
1995 *Grammatical and conceptual individuation in the lexicon*, PhD Thesis, Universiteit van Amsterdam, Ifott 15.
- Vossen, Piek and Annelies Bon  
1996 "Building a semantic hierarchy for the Sift project", Sift deliverable D20a, LRE 62030. Computer Centrum Letteren, University of Amsterdam.
- Vossen, Piek, Annelies Bon, and Ton Donker  
1996 "The semantic features in the Sift Lexicon", Sift deliverable D20b, LRE 62030. Computer Centrum Letteren, University of Amsterdam.
- Vossen, P.

- 1996 Right or wrong: combing lexical resources in the EuroWordNet project, in: M. Gellerstam, J. Jarborg, S. Malmgren, K. Noren, L. Rogstrom, C.R. Papmehl, Proceedings of Euralex-96, Goetheborg, 1996, 715-728

### **EuroWordNet Deliverables referred to**

- D000 Vossen, P.,  
1995 The EuroWordNet project, Technical Annex, EuroWordNet deliverable D000, Computer Centrum Letteren, University of Amsterdam.
- D001 Bloksma, L., P. Díez-Orzas, P. Vossen  
1996 User requirements and functional specification of the EuroWordNet project, Deliverable D001, EuroWordNet, LE2-4003, Computer Centrum Letteren, University of Amsterdam.
- D005 Climent, S., H. Rodriquez, J. Gonzalo,  
1996 Definition of the links and subsets for nouns of the EuroWordNet project, Deliverable D005, EuroWordNet, LE2-4003, Computer Centrum Letteren, University of Amsterdam.
- D006 Alonge, A.  
1996 Definition of the links and subsets for verbs, Deliverable D006, EuroWordNet, LE2-4003, Computer Centrum Letteren, University of Amsterdam.
- D007 Díez-Orzas, P., P. Forest, M. Louw  
1996 High-level Architecture of the EuroWordNet database. A Novell ConceptNet-based semantic network, Deliverable D007, EuroWordNet, LE2-4003, Computer Centrum Letteren, University of Amsterdam.

**Appendix I Nominal Base Concepts selected by all four sites**

<i>WordNet ID</i>	<i>WN Sense in</i> <i>AMS</i>	<i>Gloss</i>
00016649-n	act 1	something that people do or cause to happen
00228990-n	activity 1	"any specific activity or pursuit; they avoided all recreational activity"
09065837-n	amount of time 1	"a length of time; government services began during the colonial period"
00008030-n	animal 1	a living organism characterized by voluntary movement
00008030-n	animate being 1	a living organism characterized by voluntary movement
04111788-n	attitude 3	"a complex mental orientation involving beliefs and feelings and values and dispositions to act in certain ways; he had the attitude that work was fun"
00008030-n	beast 1	a living organism characterized by voluntary movement
05074818-n	beverage 1	any liquid suitable for drinking
00008030-n	brute 1	a living organism characterized by voluntary movement
08907331-n	chemical compound 1	(chem) a substance formed by chemical union of two or more elements or ingredients in definite proportion by weight
08805286-n	chemical element 1	"any of the more than 100 known substances (of which 93 occur naturally) that cannot be separated into simpler substances and that singly or in combination constitute all matter"
01965302-n	cloth 1	something made by weaving or felting or knitting or crocheting natural or synthetic fibers
00012878-n	cognition 1	the psychological result of perception and learning and reasoning
08907331-n	compound 4	(chem) a substance formed by chemical union of two or more elements or ingredients in definite proportion by weight
02034531-n	construction 4	a thing constructed; a complex construction or entity
00008030-n	creature 1	a living organism characterized by voluntary movement
02029323-n	decoration 2	something used to beautify
05074818-n	drink 2	any liquid suitable for drinking
05720524-n	dry land 1	the solid part of the Earth's surface
05720524-n	earth 3	the solid part of the Earth's surface
08805286-n	element 6	"any of the more than 100 known substances (of



		which 93 occur naturally) that cannot be separated into simpler substances and that singly or in combination constitute all matter"
01965302-n	fabric 1	something made by weaving or felting or knitting or crocheting natural or synthetic fibers
00008030-n	fauna 1	a living organism characterized by voluntary movement
00013522-n	feeling 1	the psychological feature of experiencing affective and emotional states
00008894-n	flora 1	a living organism lacking the power of locomotion
00011263-n	food 1	any substance that can be metabolized by an organism to give energy and build tissue
05720524-n	ground 7	the solid part of the Earth's surface
00004865-n	human 1	"a human being; there was too much for one person to do"
00016649-n	human action 1	something that people do or cause to happen
00016649-n	human activity 1	something that people do or cause to happen
00004865-n	individual 1	"a human being; there was too much for one person to do"
00012878-n	knowledge 1	the psychological result of perception and learning and reasoning
05720524-n	land 6	the solid part of the Earth's surface
08484352-n	line 26	a length (straight or curved) without breadth or thickness; the trace of a moving point
01965302-n	material 1	something made by weaving or felting or knitting or crocheting natural or synthetic fibers
08781633-n	material 5	"the tangible substance that goes into the makeup of a physical object; coal is a hard black material; wheat is the stuff they use to make bread"
00010368-n	matter 1	"that which has mass and occupies space; an atom is the smallest indivisible unit of matter"
04111788-n	mental attitude 1	"a complex mental orientation involving beliefs and feelings and values and dispositions to act in certain ways; he had the attitude that work was fun"
00004865-n	mortal 1	"a human being; there was too much for one person to do"
00011263-n	nutrient 1	any substance that can be metabolized by an organism to give energy and build tissue
02029323-n	ornament 1	something used to beautify
09065837-n	period 3	"a length of time; government services began during the colonial period"
09065837-n	period of time 1	"a length of time; government services began during the colonial period"
00004865-n	person 1	"a human being; there was too much for one person

		to do"
00019295-n	phenomenon 1	any state or process known through the senses rather than by intuition or reasoning
00008894-n	plant 1	a living organism lacking the power of locomotion
00008894-n	plant life 1	a living organism lacking the power of locomotion
05443777-n	point 12	"the precise location of something; she walked to a point where she could survey the whole street"
05074818-n	potable 1	any liquid suitable for drinking
03338771-n	quality 1	essential attribute of something or someone
05720524-n	solid ground 1	the solid part of the Earth's surface
00004865-n	someone 1	"a human being; there was too much for one person to do"
00004865-n	soul 1	"a human being; there was too much for one person to do"
02034531-n	structure 1	a thing constructed; a complex construction or entity
08781633-n	stuff 7	"the tangible substance that goes into the makeup of a physical object; coal is a hard black material; wheat is the stuff they use to make bread"
00010368-n	substance 1	"that which has mass and occupies space; an atom is the smallest indivisible unit of matter"
05720524-n	terra firma 1	the solid part of the Earth's surface
01965302-n	textile 1	something made by weaving or felting or knitting or crocheting natural or synthetic fibers
09065837-n	time period 1	"a length of time; government services began during the colonial period"
05856677-n	worker 2	a person who has employment

## Appendix II Verbal Base Concepts selected by all four sites

Total = 18 synsets

<i>WordNet ID</i>	<i>WN Sense in AMS</i>	<i>Gloss</i>
01472320-v	be 4	"copula, used with an adjective or a predicate noun: John is rich; This is not a good answer"
00432532-v	cause 6	"cause to do; cause to act in a specified manner: The ads induced me to buy a VCR; My children finally got me to buy a computer; My wife made me buy a new sofa"
00763269-v	cover 16	provide with a covering
00926361-v	create 2	"cause to be or to become; He made a mess; make friends and enemies; create a commotion"
00432532-v	get 9	"cause to do; cause to act in a specified manner: The ads induced me to buy a VCR; My children finally got me to buy a computer; My wife made me buy a new sofa"
01046072-v	go 14	"change location; move, travel, or proceed; How fast does your new car go? We travelled from Rome to Naples by bus; The policemen went from door to door looking for the suspect"
00432532-v	have 7	"cause to do; cause to act in a specified manner: The ads induced me to buy a VCR; My children finally got me to buy a computer; My wife made me buy a new sofa"
01472320-v	have the quality of being 1	"copula, used with an adjective or a predicate noun: John is rich; This is not a good answer"
00432532-v	induce 2	"cause to do; cause to act in a specified manner: The ads induced me to buy a VCR; My children finally got me to buy a computer; My wife made me buy a new sofa"
01046072-v	locomote 1	"change location; move, travel, or proceed; How fast does your new car go? We travelled from Rome to Naples by bus; The policemen went from door to door looking for the suspect"
00432532-v	make 12	"cause to do; cause to act in a specified manner: The ads induced me to buy a VCR; My children finally got me to buy a computer; My wife made me buy a new sofa"
00926361-v	make 13	"cause to be or to become; He made a mess; make friends and enemies; create a commotion"
01046072-v	move 15	"change location; move, travel, or proceed; How fast does your new car go? We travelled from Rome to Naples by bus; The policemen went from door to door looking for the suspect"
00104355-v	remove 2	"remove something concrete, as by lifting, pushing, taking off, etc.; or remove something abstract; remove a threat;

		remove a wrapper; Remove the dirty dishes from the table; take the gun from your pocket"
00432532-v	stimulate 3	"cause to do; cause to act in a specified manner: The ads induced me to buy a VCR; My children finally got me to buy a computer; My wife made me buy a new sofa"
00104355-v	take 4	"remove something concrete, as by lifting, pushing, taking off, etc.; or remove something abstract; remove a threat; remove a wrapper; Remove the dirty dishes from the table; take the gun from your pocket"
00104355-v	take away 1	"remove something concrete, as by lifting, pushing, taking off, etc.; or remove something abstract; remove a threat; remove a wrapper; Remove the dirty dishes from the table; take the gun from your pocket"
01046072-v	travel 4	"change location; move, travel, or proceed; How fast does your new car go? We travelled from Rome to Naples by bus; The policemen went from door to door looking for the suspect"

### Appendix III Top-Ontology concepts

Total number = 87

<i>Top Concept</i>	<i>Gloss</i>	<i>Opposition</i>	<i>Hyperonym</i>
Activity	all complex and planned events with a particular purpose and frequency and without a natural termination point achieved in a single event; e.g. activity, hobby, sport, education, work, performance, fight, love, caring, management		Dynamic
Animal	; e.g. animal, dog	Plant, Human, Creature	AnimateObject
AnimateObject	any object capable of living and dying, including imaginary creatures and gods; e.g. organism, life, cell, body	InanimateObject	Object
ArtActivity	all activities with the purpose to express; e.g. show, performance, dance, music		Activity
BehavioralEvent	performances of beings; e.g. do, behave, act		Dynamic
BodyPart	anything which is contained in a body; head, nose, limb, blood, finger	Not-Disjoint; can be combined with object and substance	Part
Building	; e.g. house, hotel, church, office		InanimateObject
CareEvent	all caring activities; e.g. nursing		Dynamic
CausalChange	the cause of changes; e.g. cause, prevent		Change
CausalRelation	causal relation between events or processes; e.g. cause		Relation
Change	a single change either an act or a process		Dynamic
CommunicativeEvent	all actions with the purpose of communication; e.g. speak, tell, listen, command, order, ask, state, statement, conversation, call		Dynamic

ConsumptionEvent	; e.g. eat, drink, digest		Dynamic
Container	; e.g. bag, tube, box		InanimateObject
Creature	imaginary creatures, gods; e.g. god, faust, E.T.	Animal, Human, Plant	AnimateObject
Dynamic	events or processes implying a change from one state to another state during the implied; event, act, action, become, happen, take place	Static	HighOrderEntity
EducationEvent	all activities and acts aiming at the training and learning of skills (mental and physical).		Dynamic
ExistentialChange	changes in existence; e.g. kill, produce, make, create, destroy, die, birth, die, create, destroy, delete, invent		Change
ExistentialState	all states which deal with the existence of objects and substances; e.g. exist, be, be alive, life, live, death		Static
FightActivity	all fighting activities; e.g. war		Activity
FirstOrderEntity	any concrete entity perceivable by the senses which is not a state, relation or an event; e.g. thing, anything	HighOrderEntity	Top
Functional	anything with a particular purpose but diverse denotation. Typically it can be used for nouns that can refer to any substance, object which is involved in a certain way in some event or process; e.g. remains, product, piece of art	may combine with any constitutional class	FirstOrderEntity
Furniture	; e.g. table, chair, lamp		InanimateObject
GameActivity	all activities containing an element of competition; e.g. sport, football, chess, board games, play card		Activity
Garment	; e.g. jacket, trousers, shawl		InanimateObject
Gas	cannot fall, you can inhale it and it floats above the ground; e.g. air, ozon	Liquid, Solid	Substance
Group	any concrete entity consisting of multiple discrete objects		FirstOrderEntity

	(either homogeneous or heterogenous sets), typically people, animals, vehicles; e.g. traffic, people, army, herd, fleet		
HighOrderEntity	any property, static relation or dynamic change (which cannot be grasped, heard, seen, felt as an independent physical thing); e.g. continue, occur, apply	FirstOrderEntity	Top
Human	; e.g. person, someone	Animal, Plant, Creature	AnimateObject
HumanKinship	; e.g. father, mother, brother, sister		Human
HumanNation	the people living in a Nation; e.g. American, European		Human
HumanOccupation	; e.g. doctor		Human
ImageSymbol	messages conveyed in a visual medium; e.g. sign language, traffic sign, light signal		Symbol
InanimateObject	both movable and non-movable objects; e.g. book, firm, ship, vessel, vehicle	AnimateObject	Object
Instrument	; e.g. tool, machine, weapon		InanimateObject
LanguageSymbol	messages conveyed in language (e.g. spoken, written or sign language); e.g. text, word, utterance, sentence, poem		Symbol
Liquid	can fall, feels wet and can flow on the ground; e.g. water, soup, rain	Gas, Solid	Substance
LocationState	; e.g. level, distance, separation, course, track, way, path		Static
LoveActivity	all loving activities; e.g. sex, love		Activity
ManagementEvent	all controlling activities; e.g. rule, govern, control, guide		Dynamic
Manner	the way in which an event takes place; e.g. manner, sloppy, strongly, way		HighOrderEntity
Meaning	the interpretation or message conveyed by a symbol or performance; meaning, denotation, content, topic,		Static

	story, message, interpretation		
MentalEvent	a mental state is changed; e.g. invent, remember, learn, think, consider		Dynamic
MentalState	all states that exist in mind, possibly as the result of a mental process or event; e.g. knowledge, understand, motivation, idea, fact, truth, know		Static
ModalState	states about possible situations as opposed to actual situations; e.g. ability, power, force, strength		Static
MoneySymbol	anything used for symbolizing money; e.g. share, coin		Symbol
MotionEvent	something changes location, irrespective of the causation of the change; e.g. move, put, fall, drop, drag, glide, fill, pour, empty, take out, enter		Dynamic
Object	any conceptually-countable concrete entity with an outer limit; e.g.	Substance	FirstOrderEntity
OperationalEvent	all activities or events in which an instrument is used or operated; e.g. operate, fly, drive, run		Dynamic
Part	anything which is contained in an object, substance or a group; head, juice, nose, limb, blood, finger, wheel, brick, door	can be combined with object and substance	FirstOrderEntity
Perception	mental or physical perception; notice, perceive		HighOrderEntity
Phenomenon	any state or process that occurs in nature controlled or uncontrolled; e.g. weather, chance		HighOrderEntity
PhysicalChange	changes of the physical properties of first order entities; e.g. redden, thicken, widen, enlarge, crush, form, shape, fold, wrap, thicken		Change
PhysicalColour	; e.g. colour, black, white, yellow, green, red		PhysicalState

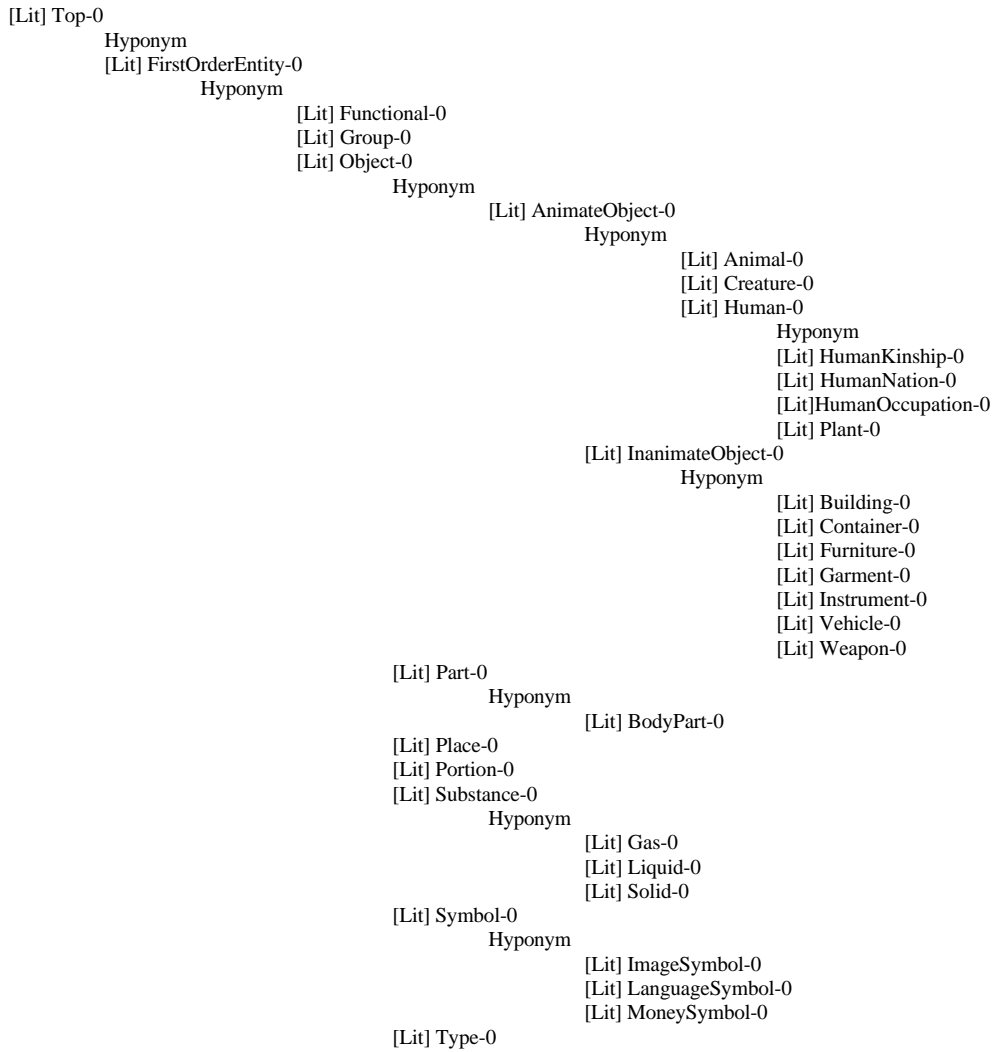


PhysicalQuantity	measure of quantity not of 3-dimensional size; e.g. couple, pair, spoonfull, meter, liter		PhysicalState
PhysicalShape	; e.g. shape, round, circular, L-spaped, spherical, square		PhysicalState
PhysicalSize	any of the 3-dimensional features of objects; e.g. length, width, depth, thickness, size		PhysicalState
PhysicalSmell	; e.g. smell, stink, odour		PhysicalState
PhysicalState	all perceptual and measurable properties of first order entities; e.g. health		Static
PhysicalTaste			PhysicalState
PhysicalTemperature	; e.g. shape, round, circular, L-spaped, spherical, square		PhysicalState
PhysicalTexture	; e.g. roughness, smoothnes		PhysicalState
PhysicalWeight	; e.g. weight, heaviness, lightness		PhysicalState
Place	concrete entities functioning as the location for something else; e.g. place, spot, centre, North, South		FirstOrderEntity
Plant	; e.g. plant, rice	Animal, Human, Creature	AnimateObject
Portion	a limited amount of a substance which can be an independent object or a part; e.g. drink, splash, piece, bite, pile, amount		FirstOrderEntity
PossessionChange	all changes in possession, often to be combined which changes in location as well; e.g. sell, buy, give, donate, steal, take, receive, send		Change
PossessionRelation	relation of possession; e.g. have, possess, possession, contain, consist of, own		Relation
QualityChange	changes of the quality of first order entities; e.g. to decline, ameliorate		Change
QualityState	e.g. deficiency		Static
QuantityChange	changes of the quantity of first order entities; e.g. to lessen, increase, decrease		Change

RecreationActivity	all activities with the purpose to recreate; e.g. vacation, fishing		Activity
Relation	a state which applies to a pair of object, states, events or processes; e.g. relation, kinship		Static
ScienceActivity	all activities with the purpose of gaining knowledge or in the field of science; e.g. biology, phylosophy		Activity
SocialState	all states relevant to a particular society; e.g. unemployment, poor, rich		Static
Solid	can fall, does not feel wet and you cannot inhale it; e.g. stone, dust, plastic, ice, metal	Liquid, Gas	Substance
SoundEvent	all audible events; e.g. song, bang, beep, rattle, snore		Dynamic
Static	all situations (properties, relations and states) in which there is no change: non-dynamic; e.g. state, property, be	Dynamic	HighOrderEntity
Substance	all stuff without boundary or fixed shape, considered from a conceptual point of view not from a linguistic point of view; e.g. mass, material	Object	FirstOrderEntity
Symbol	anything used for conveying a message; e.g. traffic sign, word, point, music		FirstOrderEntity
Time	the duration or time point of a state or event; yesterday, day, pass, long, period		HighOrderEntity
Top	all		
Type	classes, races, kinds and sorts of entities; e.g. type, race, form, class, sort, series, version, release		FirstOrderEntity
Vehicle	; e.g. car, ship, boat		InanimateObject
Weapon	; e.g. gun, pistol, magnum 48		InanimateObject
WeatherPhenomenon	any weather state or event; e.g. freeze, rain, fog, sun		Phenomenon
WorkActivity	all activities with either the purpose of making money or as		Activity

	work fulfillment in live; e.g. work, business, care		
--	--	--	--

## Appendix IV Hierarchy of EuroWordNet Top-Concepts, version 2



[Lit] HighOrderEntity-0  
 Hyponym  
     [Lit] Dynamic-0  
       Hyponym  
         [Lit] Activity-0  
           Hyponym  
             [Lit] ArtActivity-0  
             [Lit] FightActivity-0  
             [Lit] GameActivity-0  
             [Lit] LoveActivity-0  
             [Lit] RecreationActivity-0  
             [Lit] ScienceActivity-0  
             [Lit] WorkActivity-0  
         [Lit] BehavioralEvent-0  
         [Lit] CareEvent-0  
         [Lit] Change-0  
           Hyponym  
             [Lit] CausalChange-0  
             [Lit] ExistentialChange-0  
             [Lit] PhysicalChange-0  
             [Lit] PossessionChange-0  
             [Lit] QualityChange-0  
             [Lit] QuantityChange-0  
         [Lit] CommunicativeEvent-0  
         [Lit] ConsumptionEvent-0  
         [Lit] EducationEvent-0  
         [Lit] ManagementEvent-0  
         [Lit] MentalEvent-0  
         [Lit] MotionEvent-0  
         [Lit] OperationalEvent-0  
         [Lit] SoundEvent-0  
     [Lit] Manner-0  
     [Lit] Perception-0  
     [Lit] Phenomenon-0  
       Hyponym  
         [Lit] WeatherPhenomenon-0  
     [Lit] Stative-0  
       Hyponym  
         [Lit] ExistentialState-0  
         [Lit] LocationState-0  
         [Lit] Meaning-0  
         [Lit] MentalState-0  
         [Lit] ModalState-0  
         [Lit] PhysicalState-0  
           Hyponym  
             [Lit] PhysicalColour-0  
             [Lit] PhysicalQuantity-0  
             [Lit] PhysicalShape-0  
             [Lit] PhysicalSize-0  
             [Lit] PhysicalSmell-0  
             [Lit] PhysicalTemperature-0  
             [Lit] PhysicalTexture-0  
             [Lit] PhysicalWeight-0  
         [Lit] QualityState-0  
         [Lit] Relation-0  
           Hyponym  
             [Lit] CausalRelation-0  
             [Lit] PossessionRelation-0  
         [Lit] SocialState-0  
     [Lit] Time-0

## Appendix V Hierarchy of Top-Concepts with the clustering of Base Concepts, version 2

Each of the Base Concepts starts with the WordNet1.5 file offset pointer, followed by the Part of Speech tag, followed by the English Synset members or variants, followed by one or more Dutch translation. This hierarchy was created by importing the file-offset pointers as literals in the Novell ConceptNet1.0 database (Cuypers and Díez-Orzas 1996) together with a TopConcept as its hyperonym. Next all the WordNet1.5 synset members have been imported as variants and the Dutch translations as well.

```
[Lit] Top-0
  Hyponym
    [Lit] FirstOrderEntity-0
      Hyponym
        [Lit] 01958400-n-0 , thing 2-0 , zaak 1-0
          [Lit] Functional-0
            Hyponym
              [Lit] 01958716-n-0 , thing 3-0 , zaak 1-0
              [Lit] 03149538-n-0 , support 6-0 , steun 1-0
              [Lit] 02895122-n-0 , shield 2-0 , scherm 1-0
              [Lit] 05638634-n-0 , remains 2-0 , overschot 2-0
              [Lit] 02929839-n-0 , product 2-0 , production 2-0 , produkt 1-0
              [Lit] 02032220-n-0 , plaything 1-0 , toy 1-0 , speelgoed 1-0
              [Lit] 02932267-n-0 , piece of work 1-0 , work 4-0 , werk 4-0
              [Lit] 02011101-n-0 , medicament 1-0 , medication 2-0 , medicinal drug 1-0 , medicine 2-0 , middel 3-0 ,
                geneesmiddel 1-0
              [Lit] 04842062-n-0 , helping 2-0 , portion 3-0 , serving 2-0 , portie 2-0
              [Lit] 06491991-n-0 , force 7-0 , kracht 6-0
              [Lit] 06276483-n-0 , force 6-0 , power 7-0 , kracht 6-0
              [Lit] 00011263-n-0 , food 1-0 , nutrient 1-0 , voedsel 1-0
              [Lit] 05018491-n-0 , flavorer 1-0 , flavoring 1-0 , flavouwer 1-0 , flavouring 1-0 , seasoner 1-0 ,
                seasoning 2-0 , kruid 2-0
              [Lit] 02494190-n-0 , fastener 1-0 , fastening 2-0 , holdfast 1-0 , sluiting 2-0
              [Lit] 01962758-n-0 , facility 1-0 , installation 2-0 , voorziening 1-0
              [Lit] 02004554-n-0 , equipment 1-0 , materiaal 3-0 , apparatuur 1-0 , inrichting 5-0 , uitrusting 1-0
              [Lit] 02029323-n-0 , decoration 2-0 , ornament 1-0 , versiering 1-0 , versiersel 1-0
              [Lit] 05045392-n-0 , dairy product 1-0 , zuivelprodukt 1-0
              [Lit] 02024781-n-0 , curative 1-0 , cure 1-0 , remedy 2-0 , middel 3-0
              [Lit] 01992919-n-0 , creation 3-0 , produkt 1-0 , creatie 1-0
              [Lit] 01991765-n-0 , covering 4-0 , bedekking 1-0
              [Lit] 05639760-n-0 , cover 7-0 , covering 5-0 , natural covering 1-0 , bedekking 1-0 , omhulsel 1-0
              [Lit] 04858776-n-0 , confection 2-0 , confectionery 2-0 , sweet 3-0 , sweetmeat 1-0 , lekkernij 1-0
              [Lit] 02329807-n-0 , commodity 1-0 , goods 1-0 , goederen 2-0
              [Lit] 00004473-n-0 , causal agency 1-0 , causal agent 1-0 , cause 1-0 , oorzaak 1-0 , reden 1-0 ,
                veroorzaker 1-0
```

[Lit] 08128156-n-0, belonging 2-0, holding 2-0, material possession 1-0, property 5-0, bezit 1-0  
 [Lit] 02117075-n-0, barrier 1-0, afscheiding 2-0  
 [Lit] 00011607-n-0, artefact 1-0, artifact 1-0, artikel 4-0, gebruiksvoorwerp 1-0  
 [Lit] 02980374-n-0, art 2-0, fine art 1-0, kunstwerk 1-0  
 [Lit] 04837708-n-0, aliment 1-0, nourishment 1-0, nutriment 1-0, sustenance 2-0, victuals 2-0, levensmiddelen 1-0

[Lit] Group-0

Hyponym

[Lit] 01959683-n-0, unit 1-0, whole 1-0, whole thing 1-0, geheel 1-0  
 [Lit] 05354739-n-0, system 7-0, systeem 1-0, stelsel 1-0  
 [Lit] 02036726-n-0, system 1-0, unit 2-0, systeem 1-0  
 [Lit] 05142366-n-0, set 7-0, set 2-0, stel 1-0  
 [Lit] 05271053-n-0, school 5-0, school 1-0  
 [Lit] 05116476-n-0, people 1-0, volk 1-0  
 [Lit] 05149489-n-0, organization 5-0, organisatie 3-0  
 [Lit] 03553723-n-0, number 2-0, aantal 1-0  
 [Lit] 05365815-n-0, movement 7-0, richting 2-0, beweging 2-0  
 [Lit] 05334108-n-0, institute 1-0, instelling 2-0  
 [Lit] 05116306-n-0, human race 1-0, humanity 3-0, humankind 1-0, man 3-0, mankind 1-0, world 3-0, mens 1-0  
 [Lit] 00017008-n-0, group 1-0, grouping 1-0, groep 1-0, groep 4-0, groep 3-0  
 [Lit] 05152219-n-0, establishment 4-0, institution 3-0, instelling 2-0  
 [Lit] 05233198-n-0, division 9-0, afdeling 1-0  
 [Lit] 05189859-n-0, department 1-0, instituut 2-0  
 [Lit] 05218109-n-0, company 2-0, bedrijf 3-0, onderneming 2-0  
 [Lit] 05293372-n-0, commission 7-0, committee 1-0, commissie 1-0, college 2-0  
 [Lit] 05238189-n-0, club 6-0, community 3-0, fellowship 1-0, gild 1-0, guild 1-0, lodge 4-0, order 8-0, society 2-0, gemeenschap 3-0, vereniging 1-0, club 1-0, sportclub 1-0  
 [Lit] 05168576-n-0, church 3-0, faith 3-0, religion 2-0, godsdienst 1-0  
 [Lit] 05209013-n-0, body politic 1-0, commonwealth 1-0, country 2-0, land 2-0, nation 2-0, res publica 1-0, state 2-0, staat 2-0  
 [Lit] 05127029-n-0, body 7-0, instantie 1-0  
 [Lit] 05151482-n-0, authorities 1-0, government 2-0, regime 2-0, bestuur 2-0  
 [Lit] 05150995-n-0, association 3-0, instelling 2-0  
 [Lit] 05114274-n-0, arrangement 7-0, ordening 1-0  
 [Lit] 05301461-n-0, agency 1-0, authority 5-0, bureau 2-0, office 5-0, kantoor 1-0, dienst 4-0  
 [Lit] 05233375-n-0, administrative unit 1-0, dienst 4-0  
 [Lit] 05120211-n-0, accumulation 2-0, aggregation 1-0, assemblage 3-0, collection 4-0, verzameling 2-0

- [Lit] Object-0  
Hyponym
- [Lit] AnimateObject-0  
Hyponym
- [Lit] 00740781-n-0 , microorganism 1-0 , bacterie 1-0 , micro-organisme 1-0  
[Lit] 03607347-n-0 , body 3-0 , organic structure 1-0 , physical structure 1-0 ,  
lichaam 1-0  
[Lit] 00002728-n-0 , being 1-0 , life form 1-0 , living thing 1-0 ,  
organism 1-0 , organisme 2-0
- [Lit] Animal-0  
Hyponym
- [Lit] 00736689-n-0 , offspring 1-0 , young 1-0 , jong 1-0  
[Lit] 01286451-n-0 , mollusc 1-0 , mollusk 1-0 , shellfish 1-0  
[Lit] 01213903-n-0 , mammal 1-0 , zoogdier 1-0  
[Lit] 01633257-n-0 , larva 1-0 , larve 1-0  
[Lit] 01254383-n-0 , invertebrate 1-0  
[Lit] 01491542-n-0 , insect 1-0 , insekt 1-0  
[Lit] 01816356-n-0 , fish 2-0 , vis 1-0  
[Lit] 01691640-n-0 , Equus caballus 1-0 , horse 1-0 , paard 1-0  
[Lit] 01422174-n-0 , dog 1-0 , hond 1-0  
[Lit] 00884285-n-0 , bird 1-0 , vogel 1-0  
[Lit] 00008030-n-0 , animal 1-0 , animate being 1-0 , beast 1-0 ,  
brute 1-0 , creature 1-0 , fauna 1-0 ,  
wezen 1-0 , dier 1-0 , gedierte 1-0
- [Lit] Creature-0  
Hyponym
- [Lit] 05764486-n-0 , imaginary being 1-0 ,  
imaginary creature 1-0 , fabeldier 1-0  
[Lit] 05774165-n-0 , deity 1-0 , divinity 4-0 , god 2-0 ,  
immortal 1-0
- [Lit] Human-0  
Hyponym
- [Lit] 05855160-n-0 , unfortunate 1-0 , unfortunate person 1-0  
[Lit] 06313765-n-0 , ruler 2-0 , leider 1-0  
[Lit] 05853722-n-0 , religionist 1-0 , religious person 1-0 ,  
gelovige 1-0  
[Lit] 05848758-n-0 , native 1-0 , autochtoon 1-0  
[Lit] 05850734-n-0 , male 2-0 , male person 1-0 , man 1-0  
[Lit] 05850058-n-0 , leader 2-0 , leider 1-0  
[Lit] 05849094-n-0 , intellect 3-0 , intellectual 1-0 ,  
intellectueel 1-0  
[Lit] 00004865-n-0 , human 1-0 , individual 1-0 , mortal 1-0 ,  
person 1-0 , someone 1-0 , soul 1-0 ,  
persoon 1-0  
[Lit] 01779125-n-0 , homo 1-0 , human 2-0 , human being 1-0 ,  
man 1-0 , mens 1-0  
[Lit] 06102108-n-0 , friend 3-0 , vriend 1-0



- [Lit] 06093600-n-0 , follower 1-0 , aanhanger 1-0  
 [Lit] 05847495-n-0 , female 2-0 , female person 1-0 , vrouw 1-0  
 [Lit] 05846273-n-0 , expert 1-0 , deskundige 1-0 , kenner 1-0  
 [Lit] 05848227-n-0 , denizen 1-0 , dweller 1-0 , inhabitant 1-0 ,  
 inwoner 1-0  
 [Lit] 05844515-n-0 , defender 1-0 , guardian 1-0 , protector 1-0 ,  
 bewaker 1-0 , beschermer 1-0  
 [Lit] 05844200-n-0 , creator 1-0 , schepper 1-0 , maker 1-0  
 [Lit] 05852391-n-0 , compeer 1-0 , equal 2-0 , match 6-0 ,  
 peer 1-0 , gelijke 2-0  
 [Lit] 05996700-n-0 , child 1-0 , fry 1-0 , kid 2-0 , minor 1-0 ,  
 nestling 2-0 , nipper 3-0 , shaver 2-0 ,  
 small fry 1-0 , tiddler 1-0 , tike 1-0 , tyke 1-0 ,  
 youngster 1-0 , kind 1-0  
 [Lit] 06192735-n-0 , boy 3-0 , child 6-0 , male child 1-0 ,  
 jongen 1-0  
 [Lit] 05942710-n-0 , athlete 1-0 , jock 1-0 , atleet 1-0  
 [Lit] 05940574-n-0 , assistant 1-0 , help 4-0 , helper 2-0 ,  
 supporter 1-0 , knecht 1-0  
 [Lit] 05939406-n-0 , artist 1-0 , creative person 1-0 , artiest 1-0  
 [Lit] 06193747-n-0 , adult male 1-0 , man 4-0 , man 1-0  
 [Lit] 06434591-n-0 , adult female 1-0 , woman 3-0 , vrouw 1-0  
 [Lit] 06048864-n-0 , adherent 1-0 , disciple 1-0 , aanhanger 1-0  
 [Lit] 05918609-n-0 , acquaintance 2-0 , friend 1-0 , vriend 1-0  
 [Lit] 06305438-n-0 , representative 3-0 , vertegenwoordiger 1-0  
 [Lit] HumanKinship-0  
 Hyponym  
 [Lit] 06163124-n-0 , relation 3-0 ,  
 relative 2-0 , familielid 1-0  
 [Lit] 05997221-n-0 , child 2-0 , kid 3-0 ,  
 kind 2-0  
 [Lit] HumanNation-0  
 Hyponym  
 [Lit] 05873418-n-0 , European 1-0  
 [Lit] HumanOccupation-0  
 Hyponym  
 [Lit] 05856677-n-0 , worker 2-0 ,  
 arbeider 1-0  
 [Lit] 06357018-n-0 , soldier 2-0 , soldaat 1-0  
 [Lit] 06219943-n-0 , instrumentalist 1-0 ,  
 musician 2-0 , player 2-0  
 musicus 1-0 ,  
 muziekbeoefenaar 1-0  
 [Lit] 06232382-n-0 , functionary 1-0 ,  
 official 2-0 , ambtenaar 1-0  
 [Lit] 05845591-n-0 , entertainer 1-0  
 [Lit] 06069879-n-0 , employee 1-0 ,

- werknemer 1-0  
 [Lit] 06050986-n-0 , Dr. 1-0 , MD 2-0 ,  
 doc 1-0 , doctor 2-0 , medico 1-0 ,  
 physician 1-0 , arts 1-0  
 [Lit] 06438760-n-0 , author 2-0 , writer 1-0 ,  
 schrijver 1-0 , schrijver 2-0  
 [Lit] 06026990-n-0 , artificer 2-0 , artisan 1-0  
 craftsman 1-0 , journeyman 1-0 ,  
 ambachtsman 1-0  
 [Lit] 05923094-n-0 , advocate 1-0 ,  
 exponent 2-0 , proponent 1-0 ,  
 pleiter 2-0
- [Lit] Plant-0  
 Hyponym  
 [Lit] 07991027-n-0 , tree 1-0 , boom 1-0  
 [Lit] 07073185-n-0 , grass 2-0 , gras 1-0  
 [Lit] 07910410-n-0 , fungus 1-0 , zwam 1-0 , schimmel 1-0  
 [Lit] 00008894-n-0 , flora 1-0 , plant 1-0 , plant life 1-0 , gewas 1-0  
 [Lit] 07998630-n-0 , bush 4-0 , shrub 1-0 , struik 1-0
- [Lit] InanimateObject-0  
 Hyponym  
 [Lit] 02361568-n-0 , thread 1-0 , yarn 1-0 , draad 1-0 , draad 2-0 , garen 2-0  
 [Lit] 03150440-n-0 , support 7-0 , steun 1-0  
 [Lit] 02909904-n-0 , stick 3-0 , stok 3-0  
 [Lit] 03141951-n-0 , slip 9-0 , strip 4-0 , strook 1-0  
 [Lit] 02909423-n-0 , rod 3-0 , paal 1-0 , staaf 1-0  
 [Lit] 03001757-n-0 , road 2-0 , route 1-0 , weg 1-0  
 [Lit] 02908961-n-0 , pole 1-0 , paal 1-0  
 [Lit] 00009469-n-0 , inanimate object 1-0 , object 1-0 , physical object 1-0 ,  
 voorwerp 1-0  
 [Lit] 08017859-n-0 , fruit 3-0 , vrucht 1-0  
 [Lit] 03056705-n-0 , flat solid 1-0 , sheet 4-0 , blad 5-0  
 [Lit] 04935607-n-0 , edible fruit 1-0 , vrucht 2-0  
 [Lit] 02034531-n-0 , construction 4-0 , structure 1-0 , bouwsel 1-0 , bouwwerk 1-0 ,  
 constructie 3-0  
 [Lit] 02329807-n-0 , commodity 1-0 , goods 1-0 , goederen 2-0  
 [Lit] 05698341-n-0 , celestial body 1-0 , heavenly body 1-0 , hemellichaam 1-0  
 [Lit] 02245777-n-0 , card 1-0 , kaart 1-0  
 [Lit] 02174965-n-0 , book 1-0 , volume 1-0 , boek 1-0  
 [Lit] 05641227-n-0 , body 9-0 , lichaam 3-0  
 [Lit] Building-0  
 Hyponym  
 [Lit] 03066446-n-0 , shop 1-0 , store 4-0 , winkel 1-0  
 [Lit] 02725092-n-0 , room 1-0 , zaal 1-0 , vertrek 2-0 , kamer 1-0 ,  
 ruimte 3-0  
 [Lit] 01960921-n-0 , office 4-0 , kantoor 1-0  
 [Lit] 02728393-n-0 , house 2-0 , huis 1-0 , woning 1-0

- [Lit] 02895948-n-0 , factory 1-0 , manufactory 1-0 ,  
 manufacturing plant 1-0 , mill 3-0 ,  
 bedrijf 1-0 , fabriek 1-0
- [Lit] 02207842-n-0 , building 3-0 , edifice 1-0 , gebouw 1-0
- [Lit] Container-0  
 Hyponym
- [Lit] 03219464-n-0 , tube 2-0 , tubing 1-0 , buis 2-0  
 [Lit] 01990006-n-0 , container 1-0 , bak 1-0 , bergplaats 1-0  
 [Lit] 02180350-n-0 , bottle 1-0 , fles 1-0  
 [Lit] 02097669-n-0 , bag 4-0 , zak 1-0 , tas 2-0
- [Lit] Furniture-0  
 Hyponym
- [Lit] 03160216-n-0 , table 1-0 , tafel 1-0  
 [Lit] 02275608-n-0 , chair 2-0 , stoel 1-0  
 [Lit] 02008299-n-0 , article of furniture 1-0 , furniture 1-0 ,  
 piece of furniture 1-0 , kast 1-0 ,  
 meubelstuk 1-0
- [Lit] Garment-0  
 Hyponym
- [Lit] 02612319-n-0 , headdress 1-0 , headgear 1-0 , hoofddekseel 1-0  
 [Lit] 02309624-n-0 , garment 1-0 , kledingstuk 1-0  
 [Lit] 02307680-n-0 , apparel 1-0 , clothes 1-0 , clothing 1-0 ,  
 vesture 1-0 , wear 2-0 , wearing apparel 1-0 , pak 2-0 , kleding 1-0
- [Lit] Instrument-0  
 Hyponym
- [Lit] 03198235-n-0 , tool 2-0 , gereedschap 1-0 , werktuig 1-0  
 [Lit] 02804379-n-0 , musical instrument 1-0 , muziekinstrument 1-0  
 [Lit] 02010561-n-0 , mechanism 2-0 , mechanisme 2-0  
 [Lit] 02766721-n-0 , measuring instrument 1-0 , meetinstrument 1-0  
 [Lit] 02743730-n-0 , machine 2-0 , machine 1-0 , automaat 1-0  
 [Lit] 02697378-n-0 , light 1-0 , source of illumination 1-0 ,  
 lichtbron 1-0
- [Lit] 02009476-n-0 , instrumentality 1-0 , instrumentation 2-0 ,  
 installatie 3-0
- [Lit] 02657448-n-0 , instrument 2-0 , instrument 1-0  
 [Lit] 02008805-n-0 , implement 1-0 , werktuig 1-0  
 [Lit] 02004554-n-0 , equipment 1-0 , materiaal 3-0 , apparatuur 1-0 ,  
 inrichting 5-0 , uitrusting 1-0
- [Lit] 02473560-n-0 , engine 1-0 , motor 1-0  
 [Lit] 02001731-n-0 , device 2-0 , apparaat 1-0  
 [Lit] 02069513-n-0 , apparatus 1-0 , setup 2-0 , apparaat 1-0
- [Lit] Vehicle-0  
 Hyponym
- [Lit] 03233330-n-0 , vehicle 1-0 , voertuig 1-0  
 [Lit] 03061180-n-0 , ship 1-0 , schip 1-0  
 [Lit] 03235595-n-0 , craft 2-0 , vessel 1-0 , vaartuig 1-0  
 [Lit] 01991412-n-0 , conveyance 3-0 , transport 2-0 ,

- vervoermiddel 1-0  
 [Lit] 02167572-n-0 , boat 1-0 , boot 1-0  
 [Lit] 02242147-n-0 , auto 1-0 , automobile 1-0 , car 2-0 ,  
 machine 1-0 , motorcar 1-0  
 [Lit] 02051671-n-0 , aircraft 1-0 , luchtvaartuig 1-0
- [Lit] Weapon-0  
 Hyponym  
 [Lit] 03253503-n-0 , arm 4-0 , weapon 1-0 , weapon system 1-0 ,  
 wapen 1-0
- [Lit] Part-0  
 Hyponym  
 [Lit] 05467731-n-0 , surface 4-0 , vlak 1-0  
 [Lit] 02486678-n-0 , surface 1-0 , vlak 1-0  
 [Lit] 02487333-n-0 , side 1-0 , kant 1-0 , zijde 2-0  
 [Lit] 02855539-n-0 , part 3-0 , portion 2-0 , onderdeel 1-0 , deel 2-0  
 [Lit] 08450839-n-0 , part 12-0 , portion 5-0 , onderdeel 1-0  
 [Lit] 05650477-n-0 , part 10-0 , piece 10-0 , deel 2-0  
 [Lit] 02028879-n-0 , opening 4-0 , opening 3-0  
 [Lit] 05661636-n-0 , gap 4-0 , opening 11-0 , opening 3-0  
 [Lit] 08032472-n-0 , foliage 2-0 , leaf 3-0 , leafage 1-0 , blad 1-0  
 [Lit] 05382030-n-0 , face 12-0 , side 9-0 , zijde 2-0  
 [Lit] 02480168-n-0 , excavation 3-0 , hole in the ground 1-0 , kuil 1-0  
 [Lit] 05720524-n-0 , dry land 1-0 , earth 3-0 , ground 7-0 , land 6-0 , solid ground 1-0 , terra firma 1-0 ,  
 aarde 2-0 , land 1-0 , grond 6-0 , gebied 2-0  
 [Lit] 02334827-n-0 , component 1-0 , constituent 1-0 , element 1-0 , onderdeel 1-0  
 [Lit] 05383364-n-0 , bound 2-0 , boundary 2-0 , bounds 2-0 , grens 1-0  
 [Lit] 08803169-n-0 , atom 1-0 , molecule 1-0 , mote 1-0 , particle 3-0 , speck 2-0 , deeltje 1-0  
 [Lit] BodyPart-0  
 Hyponym  
 [Lit] 03734105-n-0 , vein 2-0 , vena 1-0 , ader 1-0  
 [Lit] 03632471-n-0 , tissue 1-0 , weefsel 2-0  
 [Lit] 03617358-n-0 , skin 4-0 , tegument 1-0 , huid 1-0  
 [Lit] 03622270-n-0 , passage 7-0 , passageway 2-0  
 [Lit] 03650737-n-0 , organ 4-0 , orgaan 1-0  
 [Lit] 03645458-n-0 , muscle 3-0 , musculus 1-0 , spier 1-0  
 [Lit] 03740823-n-0 , membrane 2-0 , tissue layer 1-0 , vlies 2-0  
 [Lit] 03626404-n-0 , hair 2-0 , haar 2-0 , haar 1-0  
 [Lit] 00003711-n-0 , cell 1-0 , cel 2-0  
 [Lit] 03634323-n-0 , bone 2-0 , os 1-0 , bot 2-0  
 [Lit] 03733773-n-0 , blood vessel 1-0 , bloedvat 1-0
- [Lit] Place-0  
 Hyponym  
 [Lit] 01962095-n-0 , work 3-0 , workplace 1-0 , werkplaats 1-0  
 [Lit] 02031514-n-0 , way 4-0 , weg 1-0  
 [Lit] 05450515-n-0 , region 3-0 , streek 2-0  
 [Lit] 05463659-n-0 , province 1-0 , state 5-0 , staat 2-0  
 [Lit] 05443777-n-0 , point 12-0 , punt 6-0

- [Lit] 05469653-n-0, place 13-0, spot 10-0, topographic point 1-0, gelegenheid 2-0, plaats 1-0  
 [Lit] 02857000-n-0, passage 6-0, doorgang 1-0, gang 3-0  
 [Lit] 05449837-n-0, part 9-0, region 2-0, streek 2-0  
 [Lit] 05472252-n-0, parcel 4-0, parcel of land 1-0, piece of ground 1-0, piece of land 1-0, tract 4-0, grondgebied 1-0  
 [Lit] 00014314-n-0, location 1-0, plaats 1-0  
 [Lit] 05432072-n-0, line 21-0, lijn 2-0, streep 1-0  
 [Lit] 02480168-n-0, excavation 3-0, hole in the ground 1-0, kuil 1-0  
 [Lit] 02472938-n-0, enclosure 2-0, ruimte 3-0  
 [Lit] 05657252-n-0, elevation 6-0, natural elevation 1-0, verhevenheid 2-0  
 [Lit] 05404435-n-0, district 1-0, territory 2-0, grondgebied 1-0  
 [Lit] 05657514-n-0, depression 4-0, natural depression 1-0, kom 2-0  
 [Lit] 05400698-n-0, country 3-0, land 3-0, nation 3-0, state 4-0, staat 2-0  
 [Lit] 05390395-n-0, city 2-0, metropolis 2-0, urban center 1-0, plaats 4-0, stad 1-0  
 [Lit] 05376564-n-0, area 5-0, gebied 2-0
- [Lit] Portion-0  
 Hyponym  
 [Lit] 02855539-n-0, part 3-0, portion 2-0, onderdeel 1-0, deel 2-0  
 [Lit] 02707655-n-0, layer 2-0, laag 1-0  
 [Lit] 05715416-n-0, body of water 1-0, water 5-0, water 2-0  
 [Lit] 08180701-n-0, amount 4-0, amount of money 1-0, sum 6-0, sum of money 1-0, bedrag 1-0  
 [Lit] 00018966-n-0, amount 1-0, measure 1-0, quantity 1-0, quantum 1-0, grootheid 1-0, aantal 1-0, hoeveelheid 1-0, gehalte 2-0  
 [Lit] 04141240-n-0, piece of paper 1-0, sheet 5-0, sheet of paper 1-0, vel 5-0, blad 3-0
- [Lit] Substance-0  
 Hyponym  
 [Lit] 04937211-n-0, vegetable 1-0, veggie 1-0, groente 1-0  
 [Lit] 03728455-n-0, secretion 1-0, afscheiding 3-0  
 [Lit] 05034282-n-0, sauce 1-0, saus 1-0  
 [Lit] 09018436-n-0, salt 5-0, zout 2-0  
 [Lit] 09028514-n-0, poison 2-0, vergif 1-0  
 [Lit] 09006729-n-0, pigment 1-0, verfstof 1-0  
 [Lit] 04875625-n-0, pastry 2-0, gebak 1-0, gebakje 1-0  
 [Lit] 08991530-n-0, oil 2-0, olie 1-0  
 [Lit] 08647560-n-0, neoplasm 1-0, tumor 1-0, tumour 1-0, gezwel 2-0  
 [Lit] 08783090-n-0, mixture 5-0, mengsel 1-0  
 [Lit] 04894971-n-0, meat 2-0, vlees 1-0, vleeswaren 1-0  
 [Lit] 00010368-n-0, matter 1-0, substance 1-0, stof 4-0  
 [Lit] 08781633-n-0, material 5-0, stuff 7-0, materiaal 1-0  
 [Lit] 03729776-n-0, hormone 1-0, internal secretion 1-0, hormoon 1-0  
 [Lit] 08976164-n-0, fluid 2-0, flu\i"dum 2-0  
 [Lit] 08975815-n-0, fluid 1-0, flu\i"dum 2-0  
 [Lit] 08932374-n-0, fiber 3-0, fibre 3-0, vezel 1-0  
 [Lit] 08930612-n-0, fat 3-0, vet 1-0  
 [Lit] 08918157-n-0, element 7-0, element 1-0  
 [Lit] 08919214-n-0, earth 4-0, ground 9-0, grond 6-0  
 [Lit] 02003723-n-0, drug 1-0

- [Lit] 04843172-n-0 , dish 3-0 , gerecht 1-0  
 [Lit] 04867005-n-0 , dessert 1-0 , sweet 4-0 , nagerecht 1-0  
 [Lit] 05659254-n-0 , deposit 4-0 , sediment 1-0 , afzetting 4-0  
 [Lit] 02841356-n-0 , cushioning 1-0 , padding 1-0 , vulling 1-0  
 [Lit] 08936946-n-0 , combustible 1-0 , combustible material 1-0 , fuel 1-0 , brandstof 1-0  
 [Lit] 08805286-n-0 , chemical element 1-0 , element 6-0 , element 2-0  
 [Lit] 08907331-n-0 , chemical compound 1-0 , compound 4-0 , verbinding 3-0  
 [Lit] 05050320-n-0 , cheese 1-0 , kaas 1-0  
 [Lit] 04859051-n-0 , candy 1-0 , snoepje 1-0  
 [Lit] 04879808-n-0 , cake 2-0 , koek 2-0  
 [Lit] 08885624-n-0 , building material 1-0 , bouw materiaal 1-0  
 [Lit] 04916628-n-0 , bread 1-0 , breadstuff 2-0 , staff of life 1-0 , brood 1-0  
 [Lit] 04875085-n-0 , baked good 1-0 , baked goods 1-0 , baksel 1-0  
 [Lit] 08783498-n-0 , alloy 2-0 , legering 1-0  
 [Lit] 08879673-n-0 , agent 5-0 , reagens 1-0  
 [Lit] Gas-0  
 Hyponym  
 [Lit] 08938440-n-0 , gas 5-0 , gas 1-0
- [Lit] Liquid-0  
 Hyponym  
 [Lit] 05081539-n-0 , vino 1-0 , wine 2-0 , wijn 1-0  
 [Lit] 08976498-n-0 , liquid 4-0 , vloeistof 1-0  
 [Lit] 05715416-n-0 , body of water 1-0 , water 5-0 , water 2-0  
 [Lit] 05074818-n-0 , beverage 1-0 , drink 2-0 , potable 1-0 , drank 2-0  
 [Lit] 08796177-n-0 , acid 2-0 , zuur 2-0
- [Lit] Solid-0  
 Hyponym  
 [Lit] 09057553-n-0 , wood 4-0 , hout 1-0 , houtsoort 1-0  
 [Lit] 09033134-n-0 , solid 3-0  
 [Lit] 08827122-n-0 , rock 5-0 , stone 5-0 , steen 1-0 , gesteente 1-0  
 [Lit] 05637686-n-0 , rock 4-0 , stone 2-0 , steen 1-0  
 [Lit] 08849625-n-0 , protein 1-0 , eiwit 2-0  
 [Lit] 09012321-n-0 , powder 2-0 , poeder 2-0  
 [Lit] 08996165-n-0 , paper 6-0 , papier 1-0  
 [Lit] 08983367-n-0 , mineral 1-0 , delfstof 1-0  
 [Lit] 08807415-n-0 , metal 1-0 , metallic element 1-0 , metaal 1-0  
 [Lit] 05720524-n-0 , dry land 1-0 , earth 3-0 , ground 7-0 , land 6-0 , solid ground 1-0 ,  
 terra firma 1-0 , aarde 2-0 , land 1-0 , grond 6-0 , gebied 2-0  
 [Lit] 01965302-n-0 , cloth 1-0 , fabric 1-0 , material 1-0 , textile 1-0 , kunststof 1-0 ,  
 weefsel 1-0
- [Lit] Symbol-0  
 Hyponym  
 [Lit] 04434881-n-0 , symbol 2-0 , symbool 1-0  
 [Lit] 04567799-n-0 , song 3-0 , lied 1-0  
 [Lit] 04425761-n-0 , sign 3-0 , signal 1-0 , signaling 1-0 , signaal 1-0  
 [Lit] ImageSymbol-0  
 Hyponym

[Lit] 02354709-n-0 , representation 3-0 , voorstelling 2-0 , afbeelding 1-0  
 [Lit] 02879254-n-0 , icon 1-0 , ikon 1-0 , image 2-0 , picture 1-0 , afbeelding 1-0  
 [Lit] 04481847-n-0 , emblem 2-0 , herkenningsteken 1-0  
 [Lit] 02030692-n-0 , design 2-0 , figure 2-0 , pattern 2-0 , patroon 5-0

[Lit] LanguageSymbol-0

Hyponym

[Lit] 04195435-n-0 , writing 4-0 , written material 1-0 , geschrift 1-0  
 [Lit] 04157535-n-0 , word 1-0 , woord 1-0  
 [Lit] 04183413-n-0 , title 2-0 , titel 2-0  
 [Lit] 04211005-n-0 , text 1-0 , textual matter 1-0 , tekst 1-0  
 [Lit] 08232464-n-0 , register 5-0 , register 1-0  
 [Lit] 04203578-n-0 , poem 1-0 , verse form 1-0 , gedicht 1-0  
 [Lit] 04495739-n-0 , natural language 1-0 , tongue 4-0  
 [Lit] 04180885-n-0 , name 1-0 , naam 1-0  
 [Lit] 04139704-n-0 , message 1-0 , bericht 1-0  
 [Lit] 04443464-n-0 , mark 8-0 , teken 3-0  
 [Lit] 04248202-n-0 , list 1-0 , listing 2-0 , lijst 1-0 , opsomming 1-0  
 [Lit] 04330686-n-0 , letter 1-0 , missive 1-0 , brief 1-0  
 [Lit] 04155501-n-0 , language 3-0 , linguistic communication 1-0 , taal 1-0  
 [Lit] 08225885-n-0 , document 3-0 , document 2-0 , papieren 3-0  
 [Lit] 04242515-n-0 , document 2-0 , papers 1-0 , written document 1-0  
 [Lit] 04297609-n-0 , computer program 1-0 , computer programme 1-0 , program 4-0 ,  
 programme 3-0 , computerprogramma 1-0  
 [Lit] 04444555-n-0 , character 5-0 , grapheme 1-0 , graphic symbol 1-0 , letter 1-0 ,  
 schrift 2-0  
 [Lit] 04226531-n-0 , book of facts 1-0 , reference 4-0 , reference book 1-0 ,  
 reference work 1-0 , leerboek 1-0 , naslagwerk 1-0  
 [Lit] 04222100-n-0 , book 5-0 , boek 1-0  
 [Lit] 04451043-n-0 , alphabetic character 1-0 , letter 3-0 , letter of the alphabet 1-0 ,  
 schriftteken 1-0

[Lit] MoneySymbol-0

Hyponym

[Lit] 08214427-n-0 , money 2-0 , geld 1-0  
 [Lit] 08222484-n-0 , financial obligation 1-0 , indebtedness 1-0 , liability 2-0 ,  
 pecuniary obligation 1-0 , schuld 1-0  
 [Lit] 08215253-n-0 , currency 3-0 , munteenheid 1-0  
 [Lit] 08217024-n-0 , coin 1-0 , munt 1-0

[Lit] Type-0

Hyponym

[Lit] 03957219-n-0 , form 5-0 , kind 1-0 , sort 2-0 , variety 3-0 , soort 2-0  
 [Lit] 00012356-n-0 , article 1-0 , artikel 4-0

[Lit] HighOrderEntity-0

Hyponym

[Lit] Dynamic-0

Hyponym

[Lit] 01531792-v-0 , pass 39-0 , spend 3-0 , besteden 1-0  
 [Lit] 01448761-v-0 , carry out 4-0 , do 9-0 , effect 6-0 , make 29-0

- [Lit] 01366212-v-0 , exert effort 1-0 , exert oneself 2-0 , work 21-0 , werken 1-0 , bezighouden 3-0  
 [Lit] 01341700-v-0 , act 12-0 , do something 1-0 , move 19-0 , perform an action 1-0 , take a step 2-0 ,  
 take action 1-0 , take measures 1-0 , take steps 1-0 , handelen 2-0  
 [Lit] 00980842-v-0 , do 6-0 , execute 3-0 , perform 1-0 , doen 5-0  
 [Lit] 00204516-v-0 , come about 1-0 , come to pass 1-0 , go on 1-0 , hap 2-0 , happen 1-0 , occur 1-0 ,  
 pass 16-0 , pass off 2-0 , take place 1-0 , gebeuren 2-0  
 [Lit] 00089026-v-0 , become 1-0 , come 1-0 , come out 1-0 , get 5-0 , go 4-0 , gaan 8-0 , raken 1-0  
 [Lit] 00064108-v-0 , change 11-0 , veranderen 1-0  
 [Lit] 08300433-n-0 , processing 1-0 , bewerking 1-0  
 [Lit] 08239006-n-0 , process 6-0 , proces 2-0  
 [Lit] 04690182-n-0 , happening 1-0 , natural event 1-0 , occurrence 1-0 , gebeurtenis 1-0  
 [Lit] 00597858-n-0 , group action 1-0 , actie 2-0  
 [Lit] 00016459-n-0 , event 1-0 , gebeurtenis 1-0 , evenement 1-0  
 [Lit] 08283435-n-0 , development 6-0 , evolution 2-0 , verloop 1-0  
 [Lit] 00020244-n-0 , deed 1-0 , effort 1-0 , exploit 1-0 , feat 1-0 , daad 1-0  
 [Lit] 00108829-n-0 , change 1-0 , verandering 1-0  
 [Lit] 04697176-n-0 , alteration 3-0 , change 5-0 , modification 3-0 , verandering 1-0  
 [Lit] 00021098-n-0 , action 1-0 , handeling 1-0  
 [Lit] 00016649-n-0 , act 1-0 , human action 1-0 , human activity 1-0 , handeling 1-0 , ingreep 1-0  
 [Lit] Activity-0  
 Hyponym  
 [Lit] 00228990-n-0 , activity 1-0 , bezigheid 1-0 , drukte 1-0  
 [Lit] ArtActivity-0  
 Hyponym  
 [Lit] 04326789-n-0 , show 3-0 , voorstelling 4-0  
 [Lit] 00297544-n-0 , show 1-0 , voorstelling 4-0  
 [Lit] 04487114-n-0 , performance 4-0 , public presentation 1-0 ,  
 uitvoering 3-0  
 [Lit] 04552184-n-0 , music 4-0 , muziek 2-0  
 [Lit] 00313161-n-0 , music 1-0 , muziek 2-0  
 [Lit] 00299543-n-0 , dance 1-0 , dancing 1-0 , terpsichore 1-0 ,  
 dans 1-0 , dansen 2-0  
 [Lit] 00518008-n-0 , art 1-0 , artistic creation 1-0 ,  
 artistic production 1-0 , kunst 2-0  
 [Lit] FightActivity-0  
 Hyponym  
 [Lit] 00633037-v-0 , assail 1-0 , attack 13-0 , aanvallen 2-0  
 [Lit] 00615347-v-0 , fight 5-0 , have a fight 1-0 , struggle 4-0 ,  
 strijden 1-0 , vechten 1-0  
 [Lit] 00647048-n-0 , blow 2-0 , klap 1-0  
 [Lit] 00527805-n-0 , battle 2-0 , engagement 3-0 , fight 2-0 ,  
 strijd 1-0  
 [Lit] 00540241-n-0 , attack 5-0 , onrush 1-0 , onset 1-0 ,  
 onslaught 1-0 , aanval 1-0  
 [Lit] GameActivity-0  
 Hyponym  
 [Lit] 00605818-v-0 , play 21-0 , spelen 5-0



- [Lit] 00254052-n-0 , game 1-0 , wedstrijd 1-0  
 [Lit] 04771851-n-0 , competition 3-0 , contest 2-0 , wedstrijd 1-0  
 [Lit] 00240760-n-0 , athletics 1-0 , sport 1-0
- [Lit] LoveActivity-0  
 Hyponym  
 [Lit] 00469903-n-0 , sex 1-0 , sex activity 1-0 , sexual activity 1-0 ,  
 geslachtsgemeenschap 1-0 , seks 2-0
- [Lit] RecreationActivity-0  
 [Lit] ScienceActivity-0  
 Hyponym  
 [Lit] 04037371-n-0 , science 3-0 , scientific discipline 1-0 ,  
 wetenschap 1-0 , leer 2-0  
 [Lit] 04066626-n-0 , natural philosophy 1-0 , physical science 1-0 ,  
 physics 1-0 , natuurkunde 1-0  
 [Lit] 04053427-n-0 , medical science 1-0 , medicine 3-0 ,  
 geneeskunde 1-0  
 [Lit] 04035790-n-0 , branch of knowledge 1-0 , discipline 5-0 ,  
 field 4-0 , field of study 1-0 , study 7-0 , subject 3-0 , subject area 1-0 , subject  
 field 1-0 , terrein 2-0 , vak 3-0 , discipline 2-0  
 [Lit] 04052506-n-0 , biological science 1-0 , biology 1-0 ,  
 biologie 1-0
- [Lit] WorkActivity-0  
 Hyponym  
 [Lit] 00337364-n-0 , work 1-0 , werk 1-0  
 [Lit] 00442844-n-0 , project 2-0 , task 2-0 , undertaking 1-0 ,  
 taak 1-0  
 [Lit] 00398968-n-0 , chore 1-0 , job 5-0 , task 1-0 , taak 1-0  
 [Lit] 00341191-n-0 , business 2-0 , line 1-0 , line of work 1-0 ,  
 occupation 3-0 , bezigheid 1-0  
 [Lit] 00344376-n-0 , berth 1-0 , office 1-0 , place 1-0 , position 1-0  
 post 1-0 , situation 1-0 , slot 1-0 , spot 2-0 , betrekking 1-0 , ambt 1-0
- [Lit] BehavioralEvent-0  
 Hyponym  
 [Lit] 01456537-v-0 , cozen 3-0 , deceive 2-0 , delude 1-0 , lead on 2-0 ,  
 mislead knowingly 1-0 , bedriegen 1-0  
 [Lit] 00026120-v-0 , groom 4-0 , neaten 1-0 , verzorgen 1-0  
 [Lit] 00007021-v-0 , act 7-0 , behave 1-0 , gedragen 3-0
- [Lit] CareEvent-0  
 Hyponym  
 [Lit] 01442355-v-0 , aid 6-0 , assist 4-0 , help 7-0 , helpen 1-0  
 [Lit] 00048767-v-0 , care for 1-0 , treat 2-0 , behandelen 5-0 , verzorgen 1-0  
 [Lit] 00385186-n-0 , therapy 1-0 , treatment 1-0 , behandeling 2-0 , therapie 1-0  
 [Lit] 00664219-n-0 , aid 2-0 , assistance 1-0 , help 1-0 , helping 1-0 , hulp 1-0  
 [Lit] 00383106-n-0 , aid 1-0 , attention 1-0 , care 2-0 , tending 1-0 , zorg 1-0
- [Lit] Change-0  
 Hyponym  
 [Lit] CausalChange-0

## Hyponym

- [Lit] 01432563-v-0 , assay 3-0 , attempt 4-0 , essay 3-0 , seek 3-0 ,  
try 9-0 , proberen 2-0  
[Lit] 00451248-v-0 , allow 3-0 , countenance 4-0 , let 5-0 ,  
permit 4-0  
[Lit] 01317872-v-0 , cause to have 1-0 , give 24-0  
[Lit] 00941367-v-0 , cause 7-0 , do 5-0 , give rise to 1-0 ,  
make 17-0 , veroorzaken 1-0 , aanrichten 1-0 ,  
bezorgen 2-0  
[Lit] 00432532-v-0 , cause 6-0 , get 9-0 , have 7-0 , induce 2-0 ,  
make 12-0 , stimulate 3-0 , veroorzaken 1-0

## [Lit] ExistentialChange-0

## Hyponym

- [Lit] 00926361-v-0 , create 2-0 , make 13-0 , maken 2-0  
[Lit] 00926188-v-0 , create 1-0  
[Lit] 00507790-n-0 , production 1-0 , produktie 1-0  
[Lit] 00124269-n-0 , kill 1-0 , killing 1-0 , putting to death 1-0 ,  
doding 1-0

## [Lit] PhysicalChange-0

## Hyponym

- [Lit] 00959417-v-0 , adorn 2-0 , beautify 2-0 , decorate 2-0 ,  
embellish 3-0 , grace 8-0 , ornament 2-0 , versieren 1-0  
[Lit] 00949570-v-0 , forge 6-0 , form 14-0 , mold 8-0 ,  
mould 5-0 , shape 9-0 , vormen 1-0  
[Lit] 00297919-v-0 , leave a mark on 1-0 , make a mark on 1-0  
mark 14-0 , merken 2-0  
[Lit] 00201526-v-0 , break 20-0 , come apart 1-0 , fall apart 2-0  
separate 4-0 , split up 2-0 , kapotgaan 1-0  
[Lit] 00154558-v-0 , break 19-0 , damage 6-0 , beschadigen 1-0  
[Lit] 00083270-v-0 , form 12-0 , shape 7-0 , vormen 1-0  
[Lit] 00881979-v-0 , clean 7-0 , make clean 1-0 ,  
schoonmaken 1-0 , zuiveren 1-0  
[Lit] 00778333-v-0 , connect 4-0 , connect together 1-0 ,  
link 9-0 , put together 1-0 , tie 12-0 , verbinden 1-0  
[Lit] 00768642-v-0 , fasten 3-0 , fix 11-0 , fix firmly 2-0 ,  
make fast 2-0 , secure 3-0 , vastmaken 1-0 , vastbinden 1-0  
[Lit] 00139539-n-0 , cleaning 1-0 , cleansing 1-0 , cleanup 1-0 ,  
reiniging 1-0

## [Lit] PossessionChange-0

## Hyponym

- [Lit] 01546360-v-0 , sell 7-0 , verkopen 1-0  
[Lit] 01323715-v-0 , furnish 1-0 , provide 3-0 , render 12-0 ,  
supply 6-0 , voorzien 3-0 , verschaffen 1-0  
[Lit] 01302104-v-0 , fail to profit 1-0 , lose 10-0 ,  
lose money 1-0 , make a loss 1-0 , verliezen 1-0  
[Lit] 01301401-v-0 , fail to keep 1-0 , fail to maintain 1-0 ,  
lose 8-0

- [Lit] 01281885-v-0 , make a payment 1-0 , pay 4-0 , betalen 2-0  
 [Lit] 01261345-v-0 , acquire 3-0 , get 19-0 , get hold of 3-0  
 [Lit] 01255335-v-0 , gift 4-0 , give 18-0 , give as a present 1-0 ,  
 make a present of 1-0 , present 12-0 , geven 3-0  
 [Lit] 00691086-v-0 , get hold of 2-0 , take 17-0  
 [Lit] 00671827-v-0 , cater 2-0 , ply 4-0 , provide 2-0 , supply 5-0 ,  
 voorzien 3-0  
 [Lit] 00036401-n-0 , loss 1-0 , verlies 1-0
- [Lit] QualityChange-0  
 Hyponym
- [Lit] 00122638-v-0 , decline 5-0 , get worse 1-0 , grow worse 1-0 ,  
 worsen 1-0 , verval 1-0 , achteruitgaan 2-0  
 [Lit] 00043545-v-0 , injure 1-0 , wound 4-0 , verwonden 1-0  
 [Lit] 00123997-v-0 , ameliorate 2-0 , amend 2-0 , better 6-0 ,  
 improve 2-0 , make better 1-0 , meliorate 2-0 ,  
 verbeteren 2-0
- [Lit] QuantityChange-0  
 Hyponym
- [Lit] 00262983-v-0 , decrease 6-0 , lessen 3-0 , make smaller 2-0 ,  
 minify 1-0 , verkleinen 1-0 , verminderen 2-0  
 [Lit] 00093597-v-0 , increase 7-0 , toenemen 1-0 , vermeerderen 2-0  
 [Lit] 00090574-v-0 , decrease 5-0 , diminish 1-0 , fall 11-0 ,  
 lessen 1-0 , verminderen 1-0  
 [Lit] 04725113-n-0 , increase 3-0 , toename 1-0  
 [Lit] 00204508-n-0 , increase 1-0 , step-up 1-0 , toename 1-0  
 [Lit] 00197092-n-0 , decrease 1-0 , diminution 1-0 , reduction 2-0 ,  
 step-down 1-0 , korting 1-0 , verkleining 1-0 ,  
 vermindering 1-0
- [Lit] CommunicativeEvent-0  
 Hyponym
- [Lit] 00966090-v-0 , interpret 5-0 , represent 5-0  
 [Lit] 00569629-v-0 , say 8-0 , state 7-0 , tell 7-0 , zeggen 1-0  
 [Lit] 00559904-v-0 , communicate by writing 1-0 , express by writing 1-0 , write 1-0 ,  
 schrijven 3-0  
 [Lit] 00542186-v-0 , speak 2-0 , talk 10-0 , praten 4-0  
 [Lit] 00531321-v-0 , evince 1-0 , express 6-0 , show 10-0 , uiten 2-0 , weergeven 1-0  
 [Lit] 00530290-v-0 , mouth 6-0 , speak 1-0 , talk 8-0 , utter 2-0 , verbalize 1-0 ,  
 spreken 2-0  
 [Lit] 00529407-v-0 , express 5-0 , give tongue to 1-0 , utter 1-0 , uiten 1-0  
 [Lit] 00528672-v-0 , explain 2-0 , explicate 2-0 , uitleggen 1-0  
 [Lit] 00467082-v-0 , give information 1-0 , inform 2-0 , let know 1-0 , round on 1-0 ,  
 informeren 2-0 , meedelen 2-0  
 [Lit] 00422854-v-0 , ask 1-0 , request 3-0 , vragen 3-0  
 [Lit] 00373148-v-0 , demonstrate 1-0 , establish 2-0 , prove 1-0 , shew 1-0 , show 6-0 ,  
 tonen 2-0  
 [Lit] 04388724-n-0 , statement 4-0 , beschrijving 1-0  
 [Lit] 04629714-n-0 , order 6-0 , bevel 1-0

- [Lit] 04598615-n-0 , language 5-0 , oral communication 1-0 , speech 5-0 ,  
spoken language 1-0 , taalgebruik 1-0
- [Lit] 04390828-n-0 , declaration 2-0 , verklaring 1-0
- [Lit] 04138929-n-0 , communicating 1-0 , communication 2-0 , communicatie 1-0
- [Lit] 04638292-n-0 , asking 1-0 , request 2-0 , verzoek 1-0
- [Lit] ConsumptionEvent-0  
Hyponym
- [Lit] 00656714-v-0 , consume 2-0 , have 8-0 , ingest 2-0 , take 16-0 , gebruiken 2-0
- [Lit] EducationEvent-0
- [Lit] ManagementEvent-0  
Hyponym
- [Lit] 01401683-v-0 , appoint 3-0 , charge 32-0 , aanstellen 1-0
- [Lit] 01421427-v-0 , check 28-0 , contain 3-0 , control 20-0 , curb 7-0 , hold in 2-0 ,  
moderate 7-0 , restrain 6-0 , controleren 2-0
- [Lit] 01381333-v-0 , head 28-0 , lead 21-0 , leiden 4-0
- [Lit] 01378917-v-0 , care for 4-0 , deal with 3-0 , handle 6-0 , manage 1-0 , omgaan 3-0
- [Lit] 00639819-n-0 , penalization 1-0 , penalty 1-0 , punishment 1-0 , straf 1-0
- [Lit] MentalEvent-0  
Hyponym
- [Lit] 00611702-v-0 , admit defeat 1-0 , chuck up the sponge 1-0 , drop out 1-0 ,  
give up 2-0 , leave 9-0 , quit 1-0 , throw in 2-0 , throw in the towel  
1-0 , opgeven 1-0
- [Lit] 00966090-v-0 , interpret 5-0 , represent 5-0
- [Lit] 00416049-v-0 , arrange 2-0 , order 15-0 , put 2-0 , set up 4-0 , ordenen 1-0
- [Lit] 00393722-v-0 , determine 2-0 , fix authoritatively 1-0 , fix conclusively 1-0 ,  
set 19-0 , bepalen 1-0
- [Lit] 00392710-v-0 , decide 1-0 , decide upon 1-0 , determine 1-0 ,  
make up one's mind 1-0
- [Lit] 00379073-v-0 , choose 1-0 , make a choice 1-0 , make a selection 1-0 ,  
pick out 1-0 , select 1-0 , take 10-0 , kiezen 1-0
- [Lit] 00376571-v-0 , form an opinion of 1-0 , judge 3-0 , pass judgment on 1-0 ,  
beoordelen 2-0
- [Lit] 00354465-v-0 , celebrate 1-0 , cogitate 1-0 , think 4-0 , denken 1-0
- [Lit] 00341396-v-0 , call back 1-0 , know 7-0 , recall 3-0 , recollect 1-0 , remember 1-0 ,  
remind 1-0 , retrieve 1-0 , think 2-0 , denken 3-0
- [Lit] 00396499-v-0 , devise 3-0 , plan 4-0 , think up 1-0 , bedenken 2-0
- [Lit] 03900455-n-0 , categorization 2-0 , classification 3-0 , sorting 2-0 ,  
rangschikking 1-0
- [Lit] 03885466-n-0 , act 2-0 , cognitive process 1-0 , process 3-0 , proces 2-0
- [Lit] 00091731-n-0 , choice 1-0 , selection 1-0 , keuze 2-0
- [Lit] MotionEvent-0  
Hyponym
- [Lit] 01558020-v-0 , come down 4-0 , fall 26-0 , precipitate 9-0 , vallen 1-0
- [Lit] 01323715-v-0 , furnish 1-0 , provide 3-0 , render 12-0 , supply 6-0 , voorzien 3-0 ,  
verschaffen 1-0
- [Lit] 01172741-v-0 , go by 3-0 , go past 1-0 , pass 26-0 , pass by 1-0 , travel by 1-0 ,  
voorbijgaan 1-0

- [Lit] 01152122-v-0 , come in 5-0 , enter 5-0 , get in 3-0 , get into 4-0 , go in 1-0 ,  
go into 3-0 , move into 1-0 , binnenkomen 1-0
- [Lit] 01147140-v-0 , go away 3-0 , go forth 1-0 , leave 10-0 , weggaan 1-0
- [Lit] 01144761-v-0 , arrive 1-0 , come 7-0 , get 17-0 , komen 1-0
- [Lit] 01141779-v-0 , conduct 5-0 , direct 12-0 , guide 8-0 , lead 20-0 , take 21-0
- [Lit] 01122509-v-0 , come down 3-0 , descend 1-0 , fall 15-0 , go down 3-0 , vallen 1-0 ,  
vallen 2-0 , dalen 1-0 , omvallen 1-0
- [Lit] 01114042-v-0 , ride 8-0 , rijden 2-0
- [Lit] 01084973-v-0 , foot 8-0 , go on foot 1-0 , hoof 2-0 , hoof it 1-0 , leg it 1-0 ,  
walk 11-0 , lopen 1-0
- [Lit] 01055491-v-0 , displace 3-0 , make move 1-0 , move 16-0 , bewegen 2-0 ,  
verplaatsen 1-0
- [Lit] 01054314-v-0 , depart 1-0 , go 15-0 , go away 2-0 , travel away 1-0
- [Lit] 01046072-v-0 , go 14-0 , locomote 1-0 , move 15-0 , travel 4-0 , bewegen 1-0 ,  
bewegen 6-0 , gaan 3-0
- [Lit] 01043075-v-0 , change position 1-0 , move 14-0 , bewegen 1-0
- [Lit] 00897572-v-0 , disunite 1-0 , divide 3-0 , force apart 1-0 , part 13-0 , pull apart 2-0  
separate 8-0 , take apart 2-0 , delen 2-0
- [Lit] 00867132-v-0 , project through the air 1-0 , throw 9-0 , gooien 1-0
- [Lit] 00859635-v-0 , lay 3-0 , place 23-0 , pose 4-0 , position 14-0 , put 6-0 , set 24-0 ,  
leggen 1-0 , plaatsen 1-0
- [Lit] 00827521-v-0 , bring 5-0 , bring into a certain state 1-0 , brengen 4-0
- [Lit] 00772512-v-0 , close 5-0 , shut 1-0 , afsluiten 2-0 , dichtmaken 1-0
- [Lit] 00763269-v-0 , cover 16-0 , bedekken 1-0
- [Lit] 00704074-v-0 , collide with 1-0 , hit 13-0 , hit against 1-0 , impinge on 1-0 ,  
run into 1-0 , strike 13-0 , raken 2-0
- [Lit] 00686113-v-0 , touch 18-0 , aanraken 1-0
- [Lit] 00563886-v-0 , enter 1-0 , make a record of 1-0 , put down 1-0 , record 11-0 ,  
vastleggen 3-0
- [Lit] 00268884-v-0 , fill 5-0 , fill up 2-0 , make full 1-0 , vullen 1-0
- [Lit] 00104355-v-0 , remove 2-0 , take 4-0 , take away 1-0 , verwijderen 2-0 ,  
afhalen 3-0 , losmaken 1-0 , ontdoen 1-0
- [Lit] 00329906-n-0 , stroke 3-0 , slag 3-0
- [Lit] 00064101-n-0 , push 1-0 , pushing 1-0 , duw 1-0
- [Lit] 00114479-n-0 , passage 1-0 , transition 1-0 , overgang 1-0
- [Lit] 04704743-n-0 , motion 5-0 , movement 6-0 , beweging 1-0
- [Lit] 00069655-n-0 , meeting 1-0 , touch 1-0 , touching 1-0 , aanraking 1-0
- [Lit] 00323663-n-0 , maneuver 3-0 , manoeuvre 3-0 , play 6-0 , speelwijze 1-0
- [Lit] 00172823-n-0 , journey 1-0 , journeying 1-0 , reis 1-0 , tocht 2-0
- [Lit] 00299543-n-0 , dance 1-0 , dancing 1-0 , terpsichore 1-0 , dans 1-0 , dansen 2-0
- [Lit] 00186555-n-0 , change of position 1-0 , motion 2-0 , move 5-0 , movement 2-0 ,  
beweging 1-0
- [Lit] 00157028-n-0 , change of location 1-0 , motion 1-0 , move 4-0 , movement 1-0 ,  
beweging 1-0
- [Lit] OperationalEvent-0  
Hyponym
- [Lit] 00658243-v-0 , apply 4-0 , employ 1-0 , make use of 1-0 , put to use 1-0 , use 8-0 ,

- utilise 1-0 , utilize 1-0 , gebruiken 1-0  
 [Lit] 00338477-n-0 , operation 3-0 , procedure 1-0
- [Lit] SoundEvent-0  
 Hyponym
- [Lit] 01241976-v-0 , cause to be heard 1-0 , emit sound 1-0 , go 17-0 , sound 13-0 ,  
 klinken 1-0  
 [Lit] 00554586-v-0 , emit 2-0 , express audibly 1-0 , let loose 1-0 , utter 3-0 , uiten 1-0  
 [Lit] 04599795-n-0 , vocalization 1-0 , voice 5-0 , stem 2-0  
 [Lit] 04731716-n-0 , sound 5-0 , toon 1-0 , geluid 3-0 , geluid 2-0  
 [Lit] 04567799-n-0 , song 3-0 , lied 1-0  
 [Lit] 04561287-n-0 , composition 8-0 , musical composition 1-0 , opus 1-0 , piece 7-0 ,  
 piece of music 1-0 , muziekstuk 1-0
- [Lit] Manner-0  
 Hyponym
- [Lit] 04590378-n-0 , rhetorical device 1-0 , stijlfiguur 1-0  
 [Lit] 00324581-n-0 , play 7-0 , bal 6-0  
 [Lit] 00230674-n-0 , pattern 1-0 , practice 1-0 , practise 1-0 , gewoonte 1-0  
 [Lit] 03863261-n-0 , method 2-0 , methode 1-0 , handelwijze 1-0  
 [Lit] 00096919-n-0 , means 1-0 , way 1-0 , middel 2-0 , wijze 2-0  
 [Lit] 03450012-n-0 , fashion 2-0 , manner 2-0 , mode 1-0 , style 4-0 , way 5-0 , stijl 4-0 , wijze 2-0  
 [Lit] 04575747-n-0 , expressive style 1-0 , style 7-0 , taalgebruik 1-0  
 [Lit] 08531278-n-0 , degree 6-0 , level 6-0 , point 19-0 , stage 6-0 , niveau 1-0  
 [Lit] 03540591-n-0 , degree 1-0 , grade 2-0 , level 4-0 , graad 1-0
- [Lit] Perception-0  
 Hyponym
- [Lit] 01216027-v-0 , look 8-0 , bekijken 1-0 , kijken 1-0  
 [Lit] 01203891-v-0 , experience 7-0 , get 18-0 , have 11-0 , receive 8-0 , undergo 2-0 , ervaren 2-0  
 [Lit] 01202814-v-0 , feel 12-0 , perceive 1-0 , pick up 11-0 , sense 8-0 , waarnemen 2-0  
 [Lit] 01008772-v-0 , experience 6-0 , feel 11-0 , ervaren 2-0  
 [Lit] 03858744-n-0 , sense 2-0 , sensory faculty 1-0 , sentience 2-0 , sentiency 1-0 , zintuig 1-0  
 [Lit] 03892008-n-0 , sensation 1-0 , sense datum 1-0 , sense experience 1-0 , sense impression 1-0 ,  
 gewaarwording 1-0  
 [Lit] 04792478-n-0 , pleasance 1-0 , pleasure 4-0 , plezier 1-0  
 [Lit] 00013522-n-0 , feeling 1-0 , gevoel 3-0 , neiging 1-0  
 [Lit] 04788545-n-0 , desire 2-0 , begeerte 1-0
- [Lit] Phenomenon-0  
 Hyponym
- [Lit] 00019295-n-0 , phenomenon 1-0 , verschijnsel 1-0  
 [Lit] 06464347-n-0 , natural phenomenon 1-0 , nature 5-0 , natuurverschijnsel 1-0  
 [Lit] 06467144-n-0 , chance 3-0 , fortune 1-0 , hazard 2-0 , luck 1-0 , kans 1-0  
 [Lit] 04701573-n-0 , bad luck 1-0 , misfortune 1-0 , ongeluk 1-0  
 [Lit] WeatherPhenomenon-0  
 Hyponym
- [Lit] 06529752-n-0 , wind 7-0 , wind 1-0  
 [Lit] 06529389-n-0 , atmospheric condition 1-0 , elements 1-0 , weather 1-0 , weer 2-0
- [Lit] Stative-0  
 Hyponym

- [Lit] 01515519-v-0 , hold 26-0 , keep 16-0 , maintain 9-0 , handhaven 1-0 , houden 9-0  
 [Lit] 01503041-v-0 , agree 5-0 , correspond 3-0 , fit 9-0 , gibe 3-0 , jibe 3-0 , match 14-0 , tally 7-0 ,  
 overeenkomen 1-0  
 [Lit] 01472320-v-0 , be 4-0 , have the quality of being 1-0 , zijn 7-0  
 [Lit] 03564110-n-0 , value 2-0 , waarde 2-0 , waarde 1-0  
 [Lit] 00015437-n-0 , state 1-0 , toestand 1-0  
 [Lit] 08522741-n-0 , situation 4-0 , state of affairs 1-0 , toestand 1-0  
 [Lit] 03338771-n-0 , quality 1-0 , eigenschap 1-0  
 [Lit] 03444246-n-0 , property 2-0 , eigenschap 1-0  
 [Lit] 08530753-n-0 , natural state 1-0 , nature 6-0 , state of nature 1-0 , wild 2-0 , natuur 1-0  
 [Lit] 00017586-n-0 , attribute 1-0 , eigenschap 1-0  
 [Lit] ExistentialState-0  
     Hyponym  
       [Lit] 01471536-v-0 , be 3-0 , exist 1-0 , zijn 2-0  
 [Lit] LocationState-0  
     Hyponym  
       [Lit] 01501697-v-0 , be 9-0 , occupy a certain area 1-0 , occupy a certain position 1-0 ,  
         bevinden 2-0 , bevinden 3-0 , zijn 3-0  
       [Lit] 01492762-v-0 , stay in one place 1-0 , wait 5-0 , wachten 1-0  
       [Lit] 00685874-v-0 , adjoin 1-0 , meet 4-0 , touch 17-0 , aanraken 1-0  
       [Lit] 00015245-n-0 , space 1-0 , ruimte 3-0  
       [Lit] 08434357-n-0 , elbow room 1-0 , room 3-0 , way 10-0 , ruimte 1-0  
       [Lit] 03536009-n-0 , distance 1-0 , afstand 1-0  
       [Lit] 08463109-n-0 , direction 8-0 , richting 1-0  
       [Lit] 05477069-n-0 , direction 7-0 , way 8-0 , richting 1-0  
 [Lit] Meaning-0  
     Hyponym  
       [Lit] 01513147-v-0 , be about 2-0 , bear on 3-0 , come to 6-0 , concern 6-0 ,  
         have to do with 1-0 , pertain 2-0 , refer 2-0 , relate 5-0 , touch 25-0 , touch on 6-0 , gaan 1-0  
       [Lit] 00537777-v-0 , intend 4-0 , mean 6-0 , betekenen 1-0  
       [Lit] 04337839-n-0 , info 1-0 , information 2-0 , informatie 1-0  
       [Lit] 04430266-n-0 , indication 1-0 , indicatie 1-0  
       [Lit] 03953834-n-0 , idea 2-0 , thought 2-0 , gedachte 2-0 , idee 1-0  
       [Lit] 04313427-n-0 , content 3-0 , message 2-0 , subject matter 1-0 , substance 4-0 ,  
         stof 6-0  
       [Lit] 00018599-n-0 , communication 1-0  
 [Lit] MentalState-0  
     Hyponym  
       [Lit] 01040073-v-0 , desire 4-0 , want 8-0 , willen 1-0  
       [Lit] 00452960-v-0 , agree 2-0 , concur 2-0 , goedkeuren 2-0 , instemmen 1-0  
       [Lit] 00405636-v-0 , await 1-0 , expect 4-0 , look 6-0 , look forward to 1-0 , wait 3-0 ,  
         wait for 1-0 , verwachten 1-0  
       [Lit] 00388394-v-0 , consider 1-0 , reckon 3-0 , regard 7-0 , see 5-0 , take to be 2-0 ,  
         view 11-0 , beschouwen 2-0  
       [Lit] 00387631-v-0 , believe 3-0 , think 6-0 , denken 4-0  
       [Lit] 00333362-v-0 , cognize 1-0 , know 3-0 , weten 2-0  
       [Lit] 03282629-n-0 , trait 1-0 , neiging 1-0

- [Lit] 04033925-n-0 , theory 3-0 , theorie 2-0  
 [Lit] 03864615-n-0 , system 4-0 , system of rules 1-0 , systeem 1-0  
 [Lit] 03898550-n-0 , structure 4-0 , structuur 2-0  
 [Lit] 00566905-n-0 , procedure 3-0 , process 1-0 , proces 2-0 , werkwijze 1-0  
 [Lit] 04792478-n-0 , pleasance 1-0 , pleasure 4-0 , plezier 1-0  
 [Lit] 03985547-n-0 , plan 3-0 , program 1-0 , programme 1-0 , plan 2-0 , plan 1-0  
 [Lit] 00013299-n-0 , motivation 1-0 , motive 1-0 , need 1-0 , reden 1-0  
 [Lit] 03944302-n-0 , information 1-0 , informatie 1-0  
 [Lit] 04827440-n-0 , humor 3-0 , mood 1-0 , temper 4-0 , stemming 1-0  
 [Lit] 04011318-n-0 , faith 2-0 , religion 1-0 , religious belief 1-0 , godsdienst 1-0  
 [Lit] 04785784-n-0 , emotion 1-0 , emotie 1-0  
 [Lit] 04009596-n-0 , doctrine 1-0 , ism 1-0 , philosophy 1-0 , school of thought 1-0 ,  
 filosofie 2-0  
 [Lit] 04113320-n-0 , disposition 4-0 , inclination 4-0 , tendency 2-0 , neiging 1-0  
 [Lit] 04788545-n-0 , desire 2-0 , begeerte 1-0  
 [Lit] 03954891-n-0 , concept 1-0 , conception 3-0 , concept 2-0  
 [Lit] 00012878-n-0 , cognition 1-0 , knowledge 1-0 , kennis 2-0  
 [Lit] 03433579-n-0 , behavior 3-0 , behaviour 3-0 , conduct 2-0 , demeanor 1-0 ,  
 demeanour 1-0 , deportment 1-0 , gedrag 1-0  
 [Lit] 04111788-n-0 , attitude 3-0 , mental attitude 1-0 , houding 2-0  
 [Lit] 03898749-n-0 , arrangement 4-0 , organization 4-0 , system 5-0 , organisatie 2-0 ,  
 systeem 1-0  
 [Lit] 08549511-n-0 , accord 4-0 , agreement 6-0 , overeenkomst 2-0
- [Lit] ModalState-0  
 Hyponym
- [Lit] 01217877-v-0 , appear 6-0 , look 9-0 , seem 1-0 , lijken 2-0  
 [Lit] 01539155-v-0 , can 8-0 , kunnen 2-0  
 [Lit] 03854243-n-0 , inability 2-0 , onvermogen 1-0  
 [Lit] 03857413-n-0 , faculty 1-0 , mental faculty 1-0 , module 4-0 , vermogen 3-0  
 [Lit] 03849803-n-0 , accomplishment 2-0 , acquirement 1-0 , acquisition 2-0 ,  
 attainment 3-0 , skill 2-0 , techniek 3-0 , vaardigheid 1-0  
 [Lit] 03841132-n-0 , ability 2-0 , power 3-0 , vermogen 3-0  
 [Lit] 03601639-n-0 , ability 1-0 , vermogen 3-0
- [Lit] PhysicalState-0  
 Hyponym
- [Lit] PhysicalColour-0  
 Hyponym
- [Lit] 03463765-n-0 , color 2-0 , coloring 2-0 , colour 2-0 ,  
 colouring 2-0 , kleur 1-0
- [Lit] PhysicalQuantity-0  
 Hyponym
- [Lit] 08313335-n-0 , unit 6-0 , unit of measurement 1-0, eenheid 2-0  
 [Lit] 03966324-n-0 , quantity 3-0 , groetheid 1-0  
 [Lit] 08317731-n-0 , number 10-0 , getal 1-0  
 [Lit] 03539714-n-0 , measure 5-0 , measurement 2-0 , graad 1-0
- [Lit] PhysicalShape-0  
 Hyponym



- [Lit] 08482581-n-0 , solid 1-0 , lichaam 3-0  
 [Lit] 08484352-n-0 , line 26-0 , lijn 2-0 , streep 1-0  
 [Lit] 00014558-n-0 , form 1-0 , shape 1-0 , vorm 1-0  
 [Lit] 08483587-n-0 , figure 12-0 , figuur 4-0  
 [Lit] 03314728-n-0 , appearance 4-0 , visual aspect 1-0 , uiterlijk 1-0
- [Lit] PhysicalSize-0  
 [Lit] PhysicalSmell-0  
 [Lit] PhysicalTemperature-0  
 [Lit] PhysicalTexture-0  
 [Lit] PhysicalWeight-0
- [Lit] QualityState-0  
 Hyponym
- [Lit] 08658681-n-0 , plant disease 1-0 , planteziekte 1-0  
 [Lit] 08587853-n-0 , illness 1-0 , malady 1-0 , sickness 1-0 , ziekte 1-0  
 [Lit] 08665752-n-0 , harm 3-0 , hurt 5-0 , injury 4-0 , trauma 2-0 , verwonding 2-0  
 [Lit] 08682700-n-0 , flush 4-0 , inflammation 1-0 , redness 2-0 , ontsteking 1-0  
 [Lit] 08693431-n-0 , disturbance 7-0 , mental disorder 1-0 , mental disturbance 1-0 ,  
 stoornis 2-0  
 [Lit] 08586618-n-0 , disorder 2-0 , upset 4-0 , gebrek 4-0 , stoornis 2-0 , aandoening 1-0  
 [Lit] 08592183-n-0 , disease 1-0 , ziekte 1-0  
 [Lit] 08731035-n-0 , deficiency 2-0 , lack 1-0 , want 2-0 , gebrek 2-0  
 [Lit] 08738373-n-0 , defect 3-0 , fault 4-0 , flaw 2-0 , fout 2-0 , onvolkomenheid 1-0  
 [Lit] 08520394-n-0 , condition 5-0 , status 2-0 , gesteldheid 1-0 , hoedanigheid 1-0  
 [Lit] 08520221-n-0 , condition 4-0 , voorwaarde 2-0
- [Lit] Relation-0  
 Hyponym
- [Lit] 01475351-v-0 , end 16-0 , finish 11-0 , terminate 4-0 , eindigen 2-0  
 [Lit] 01506899-v-0 , be 10-0 , be identical to 2-0 , equal 4-0  
 [Lit] 08711637-n-0 , union 9-0 , vereniging 2-0  
 [Lit] 08523811-n-0 , relationship 4-0 , hoedanigheid 2-0  
 [Lit] 08523567-n-0 , relationship 3-0 , hoedanigheid 2-0  
 [Lit] 00017862-n-0 , relation 1-0 , relatie 1-0 , relatie 3-0  
 [Lit] 08457189-n-0 , ratio 1-0 , verhouding 1-0  
 [Lit] 08453309-n-0 , family relationship 1-0 , kinship 2-0 , relationship 2-0 ,  
 hoedanigheid 2-0  
 [Lit] 08440487-n-0 , connectedness 1-0 , connection 7-0 , connexion 5-0 , verband 2-0
- [Lit] CausalRelation-0  
 Hyponym
- [Lit] 04349399-n-0 , policy 3-0 , beleid 1-0  
 [Lit] 00038929-n-0 , error 1-0 , fault 1-0 , mistake 1-0 , fout 2-0  
 [Lit] 06465491-n-0 , consequence 3-0 , effect 4-0 , outcome 2-0 ,  
 result 3-0 , upshot 1-0 , werking 2-0 , effect 1-0  
 , gevolg 1-0  
 [Lit] 04029556-n-0 , aim 2-0 , object 3-0 , objective 2-0 , target 2-0  
 doel 3-0
- [Lit] PossessionRelation-0  
 Hyponym

- [Lit] 01257491-v-0 , have 13-0 , have possession of 1-0 , own 1-0 ,  
possess 1-0 , hebben 1-0
- [Lit] 01256853-v-0 , have 12-0 , have got 1-0 , hold 19-0
- [Lit] 00017394-n-0 , possession 1-0 , bezit 1-0
- [Lit] 08179398-n-0 , asset 2-0 , bezit 1-0
- [Lit] SocialState-0  
Hyponym
- [Lit] 03586387-n-0 , right 4-0 , recht 4-0
- [Lit] 08534455-n-0 , position 13-0 , status 3-0 , positie 6-0
- [Lit] 00342842-n-0 , employment 1-0 , job 3-0 , work 2-0 , baan 1-0
- [Lit] 08550427-n-0 , disorder 1-0 , chaos 1-0
- [Lit] 08719491-n-0 , dignity 3-0 , waardigheid 2-0
- [Lit] 08535290-n-0 , degree 7-0 , rank 2-0 , social rank 1-0 , social station 1-0 ,  
social status 1-0 , onderscheidingsteken 1-0
- [Lit] Time-0  
Hyponym
- [Lit] 01517254-v-0 , continue 7-0 , go on 6-0 , keep 17-0 , keep on 2-0 , proceed 5-0 , houden 8-0
- [Lit] 00210630-v-0 , continue 2-0 , go along 1-0 , go on 2-0 , persist in 1-0 , doorgaan 3-0
- [Lit] 00068138-v-0 , continue 1-0 , remain 1-0 , rest 9-0 , stay 7-0 , blijven 5-0
- [Lit] 01515268-v-0 , cease 3-0 , discontinue 2-0 , give up 12-0 , lay off 2-0 , quit 5-0 , stop 20-0 ,  
opgeven 1-0
- [Lit] 00218979-v-0 , break 23-0 , break off 2-0 , discontinue 1-0 , stop 13-0 , stoppen 5-0
- [Lit] 00213455-v-0 , bring to an end 1-0 , cause to end 1-0 , end 15-0 , terminate 2-0 , beeindigen 1-0
- [Lit] 09077332-n-0 , clock time 1-0 , time 6-0 , tijdeenheid 1-0
- [Lit] 00014882-n-0 , time 1-0 , tijd 1-0
- [Lit] 09157756-n-0 , instant 1-0 , minute 5-0 , moment 5-0 , second 9-0 , tijdstip 1-0
- [Lit] 09094193-n-0 , day 5-0 , dag 3-0
- [Lit] 09131680-n-0 , calendar month 1-0 , month 2-0 , maand 1-0
- [Lit] 09065837-n-0 , amount of time 1-0 , period 3-0 , period of time 1-0 , time period 1-0 , tijd 3-0 ,  
periode 1-0

**Appendix VI Genuine Drop-Outs AMS**

<b>TopConcept</b>	<b>AMS Genuine Drop Outs, proposed for re-selection as BCs</b>
Group	05127861-n, maatschappij 1, society 1
AnimateObject	00511687-n, cultuur 3, culture 2
Dynamic	00071241-v, veranderen 4, alter 2, change 12, vary 1
FightActivity	00631049-v, verzetten 2, hold out 2, offer resistance 1, resist 2, stand firm 2
RecreationActivity	05255204-n, feest 1, party 2, 04769704-n, party 1
BehavioralEvent	00448333-v, weigeren 2, decline 7, refuse 2
	01511883-v, opvallen 1, excel 1, stand out 3, surpass 3
	04412425-n, grap 1, crack 6, quip 1, sally 3, wisecrack 1
	00422340-n, fraude 1, dupery 1, fraud 1, fraudulence 1, hoax 1
	01422880-v, benadelen 1, disadvantage 3, disfavor 3, disfavour 3
CausalChange	01431735-v, mislukken 1, fail 7, go wrong 2, miscarry 2
ExistentialChange	04711277-n, ontstaan 2, genesis 1, inception 1, origin 3, origination 2
	00216283-v, doodgaan 1, sterven 1, decease 2, die 6, exit 4
PhysicalChange	00223859-v, verwarmen 1, make warm 1, warm 1
	00237247-v, uiteenvallen 1, break into parts 1, disintegrate 1
	00772933-v, openen 3, cause to open 1, open 8, open up 4
	00302695-v, bewerken 1, process 7, treat 3
PossessionChange	01277199-v, verkopen 2, exchange for money 1, sell 4
	00430622-v, verkopen 2, sell 2
	01260836-v, krijgen 1, have 15, receive 9
	00307705-v, krijgen 1, find 3, get 6, obtain 1, receive 2
	01259481-v, kopen 1, buy 3, purchase 3, take 25
	00434296-n, diefstal 1, larceny 1, stealing 1, theft 1, thievery 1
	01320281-v, afnemen 6, rob 2
	01318941-v, afnemen 6, fleece 6, gazump 2, hook 17, overcharge 3
QualityChange	00928634-v, kapotmaken 1, destroy 4
	00900879-v, kapotmaken 1, vernielen 1, vernietigen 1, destroy 3, ruin 6
QuantityChange	00796914-v, verzamelen 2, accumulate 2, bring together 3, collect 3, get together 2
	00794237-v, verzamelen 2, collect 2, garner 2, gather 7
CommunicativeEvent	00273772-v, zwijgen 1, become quiet 1, become still 1, fall silent 1, hush 4
	00581359-v, noemen 1, call 20, name 11
	00547315-v, noemen 1, call 19
EducationEvent	03911038-n, opleiding 1, education 4
	00405488-n, opgave 1, school assignment 1
	00491582-n, onderwijs 1, opleiding 1, scholing 1, education 2,

	educational activity 1
	00498414-n, oefening 2, drill 1, exercise 3, practice 4, practice session 1
	00368355-n, oefening 1, exercise 1, exercising 1, physical exertion 1, workout 1
ManagementEvent	00361645-n, politiek 1, politics 1
	00102660-n, verkiezingen 1, election 2
	00101987-n, verkiezingen 1, election 1
	00372646-n, rechtspraak 1, law practice 1
	05289006-n, overleg 1, get together 1, meeting 6
	05288049-n, overleg 1, vergadering 1, meeting 5
MentalEvent	00442867-v, onderzoeken 2, investigate 2, look into 2
	01274026-v, weigeren 3, decline 10, pass up 2, refuse 4, reject 5
	00351672-v, lezen 1, read 6
	00375861-n, controle 1, onderzoek 1, examination 2, inspection 1, scrutiny 1
	01436060-v, "beïnvloeden 1, act upon 2, exert influence upon 1, influence 8
MotionEvent	00796914-v, verzamelen 2, accumulate 2, bring together 3, collect 3, get together 2
	00794237-v, verzamelen 2, collect 2, garner 2, gather 7
	05347181-n, verkeer 1, traffic 3
	00253904-v, verdwijnen 1, disappear 1, go away 1, vanish 2
	01224184-v, verbergen 1, verbergen 3, conceal 1, hide 5
	00700678-v, steken 3, dig 6, jab 3, poke 5, prod 3
	01279432-v, halen 1, bring 10, bring in 4, fetch 3, go for 6
	00678466-n, bezoek 1, visit 3
	01447000-v, belemmeren 1, block 21, blockade 6, embarrass 2, hinder 3
Perception	08677522-n, pijn 1, hurting 1, pain 4
	03897733-n, pijn 1, pain 1, painful sensation 1
	03892442-n, geur 1, odor 2, odour 2, olfactory perception 1, olfactory sensation 1
	03476724-n, geur 1, aroma 1, odor 1, odour 1, olfactory property 1
Phenomenon	06487051-n, stroom 3, electric power 1, electricity 3
	06502153-n, licht 10, light 12, visible light 1, visible radiation 1
Static	08548403-n, rust 3, quiet 3, tranquillity 2
	03437297-n, rust 3, placidity 1, quiet 1, repose 2, serenity 1
ExistentialState	01479650-v, leven 8, live 4
	01513610-v, gelden 3, apply 10, go for 7, hold 25
LocationState	05441398-n, route 1, path 3, route 2
MentalState	08479600-n, norm 2
	04678120-n, norm 2, criterion 2, rule 8, standard 3, touchstone 1

	103956560-n, voorstelling 1, idea 3
	06359978-n, ziel 1, psyche 2, soul 4
	00406425-v, vertrouwen 2, believe 4, trust 9
	00386671-v, vertrouwen 2, bank on 1, confide in 1, have confidence in 1, have faith in 1
	04009451-n, vertrouwen 1, faith 1, trust 3
	00009337-v, slapen 1, catch some Z's 1, kip 2, log Z's 1, sleep 4
	04031912-n, cultuur 2, acculturation 3, culture 4
	03938080-n, besef 1, apprehension 2, discernment 5, savvy 1, understanding 2
	03871589-n, besef 1, awareness 1, cognizance 1, knowingness 2
	00330150-v, begrijpen 1, understand 1
ModalState	03859305-n, wil 1, volition 2, will 1
	08744158-n, mogelijkheid 1, possibility 4, possibleness 1
	04014029-n, mogelijkheid 1, possibility 3
	03930751-n, mogelijkheid 1, opening 7, possibility 1, possible action 1
	04421357-n, moeilijkheid 1, problem 2
	03877957-n, moeilijkheid 1, probleem 1, problem 1, trouble 2
PhysicalSize	03542970-n, maat 6, size 2
	03563097-n, hoogte 1, height 3
	03488815-n, hoogte 1, height 1, stature 1
PhysicalTemperat.	03497241-n, warmte 1, warmth 1, warmth 3
PhysicalWeight	03503480-n, gewicht 1, weight 3
QualityState	04795555-n, ongemak 1, discomfort 1, irritation 3, soreness 1, tenderness 2
Relation	03352335-n, verschil 1, difference 1
CausalRelation	04414253-n, voorwaarde 1, condition 3, term 4
	04694624-n, resultaat 1, uitslag 2, outcome 1, result 2
	03864476-n, oplossing 4, solution 1
	08440293-n, grondslag 2, basis 2, footing 2, ground 8
PossessionRelation	01537537-v, dragen 3, carry 27, pack 22, take 36
SocialState	05277423-n, cultuur 2, culture 5
	03908336-n, cultuur 2, culture 3
	05358864-n, wet 1, jurisprudence 2, law 7
	04281206-n, wet 1, law 5
Time	09168906-n, begin 1, beginning 5, commencement 6, first 6, kickoff 2
	00207968-v, beginnen 6, begin 1, commence 1, set about 1, set out 1

**Appendix VII FUE-Sample of the Noun database (hyponymy)**

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n {animal,bestia,criatura,fauna,ser\_animado} 00736689-n {cría,vástago}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008894-n {flora,planta,vida\_vegetal} 07910410-n {hongo}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008894-n {flora,planta,vida\_vegetal} 06550520-n {anual,planta\_anual}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n {animal,bestia,criatura,fauna,ser\_animado} 00849436-n {cordado} 00854210-n {vertebrado} 00884285-n {ave}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n {animal,bestia,criatura,fauna,ser\_animado} 00733894-n {bicho,sabandija}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00775750-n {parásito}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 06220762-n {mutación,mutante}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 06148720-n {individuo}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n {animal,bestia,criatura,fauna,ser\_animado} 00738439-n {gigante}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008894-n {flora,planta,vida\_vegetal} 07974178-n {planta\_vascular} 07990292-n {planta\_leñosa} 07998630-n {arbusto,mata}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n {animal,bestia,criatura,fauna,ser\_animado} 00844811-n {embrión}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00741291-n {cruce,híbrido}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n {animal,bestia,criatura,fauna,ser\_animado} 01254383-n {invertebrado} 01286451-n {molusco}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00774801-n {plancton}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n {animal,bestia,criatura,fauna,ser\_animado} 00736126-n {hembra}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00740781-n {microorganismo}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008894-n {flora,planta,vida\_vegetal} 07974178-n {planta\_vascular} 07169764-n {planta\_herbácea} 07072915-n {gramínea} 07073185-n {hierba}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n {animal,bestia,criatura,fauna,ser\_animado} 01698815-n {corredor}

00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n  
 {animal,bestia,criatura,fauna,ser\_animado} 01254383-n {invertebrado} 01126858-n  
 {artrópodo} 01491542-n {insecto}  
 00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n  
 {animal,bestia,criatura,fauna,ser\_animado} 00736620-n {adulto}  
 00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n  
 {animal,bestia,criatura,fauna,ser\_animado} 01633257-n {larva}  
 00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00003504-n  
 {vida} 05140379-n {fauna}  
 00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n  
 {animal,bestia,criatura,fauna,ser\_animado} 06367426-n {pareja}  
 00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n  
 {animal,bestia,criatura,fauna,ser\_animado} 00849436-n {cordado} 00854210-n  
 {vertebrado} 00855637-n {vertebrado\_acuático} 01816356-n {pez}  
 00002403-n {entidad} 00002728-n {forma\_de\_vida,organismo,ser,ser\_vivo} 00008030-n  
 {animal,bestia,criatura,fauna,ser\_animado} 01485573-n {presa}  
 00002403-n {entidad} 00003711-n {célula} 03757681-n {célula\_nerviosa,neurona}

**Appendix VIII FUE-Sample of the Verb database (hyponymy)**

00064108-v {cambiar,sufrir\_un\_cambio} 00081466-v {cambiar\_fisicamente} 00201526-v {partirse,romperse}  
00064108-v {cambiar,sufrir\_un\_cambio} 00086015-v {convertirse,transformarse}  
00089026-v {devenir,llegar\_a\_ser,ponerse,volverse}  
00064108-v {cambiar,sufrir\_un\_cambio} 00086015-v {convertirse,transformarse}  
00122638-v {decaer,empeorar}  
00064108-v {cambiar,sufrir\_un\_cambio} 00101800-v {cambiar\_de\_magnitud}  
00090574-v {disminuir,reducirse}  
00064108-v {cambiar,sufrir\_un\_cambio} 00101800-v {cambiar\_de\_magnitud}  
00093597-v {aumentar}  
00064108-v {cambiar,sufrir\_un\_cambio} 00154558-v {dañar,estropear,romper}  
00072540-v {alterar,cambiar} 00040824-v {indisponer} 00040663-v {dañar,hacer\_daño}  
00043545-v {herir}  
00072540-v {alterar,cambiar} 00083270-v {dar\_forma}  
00072540-v {alterar,cambiar} 00091455-v {acrecetar,incrementar}  
00072540-v {alterar,cambiar} 00123997-v {mejorar}  
00072540-v {alterar,cambiar} 00213455-v {acabar,finalizar,terminar} 00218979-v {interrumpir,terminar}  
00072540-v {alterar,cambiar} 00262983-v {disminuir,reducir}  
00072540-v {alterar,cambiar} 00268884-v {llenar}  
00072540-v {alterar,cambiar} 00297919-v {marcar,señalar}  
00072540-v {alterar,cambiar} 00827521-v {causar,producir,provocar}  
00072540-v {alterar,cambiar} 00881979-v {limpiar}  
00072540-v {alterar,cambiar} 00959417-v {adornar,decorar,embellecer,ornar}  
00354465-v {discurrir,pensar} 00396499-v {pensar,planear}  
00374169-v {confirmar,corroborar} 00373148-v {demostrar,probar}  
00376571-v {juzgar,opinar} 00387631-v {considerar,creer,pensar} 00388394-v {considerar,ver}  
00392710-v {decidir,determinar} 00379073-v {coger,escoger,seleccionar}  
00392710-v {decidir,determinar} 00392562-v {resolver} 00393722-v {concluir,determinar,resolver} 00348034-v {identificar} 00365740-v {diferenciar,distinguir}  
00529407-v {decir\_verbalmente,expresar} 00569629-v {afirmar,decir,manifestar}  
00570287-v {declarar}  
00605050-v {competir,contender,medirse} 00605818-v {jugar}  
00615347-v {luchar,pelearse} 00633037-v {asaltar,atacar}  
00686113-v {contactar,tocar} 00704074-v {golpear}  
00778333-v {unir} 00743265-v {sujetar} 00768642-v {asegurar,fijar}  
00926361-v {causar,crear,hacer,realizar} 00928226-v {recrear} 00966090-v {representar}  
00926361-v {causar,crear,hacer,realizar} 00928226-v {recrear} 00980842-v {ejecutar,hacer,llevar\_a\_cabo}



00926361-v {causar,crear,hacer,realizar} 00941367-v {hacer}  
 00926361-v {causar,crear,hacer,realizar} 00945714-v {elaborar,fabricar} 00949570-v  
 {dar\_forma,formar}  
 00926361-v {causar,crear,hacer,realizar} 01003070-v {despertar,estimular,incitar}  
 01004175-v {conmover}  
 01043075-v {moveuse} 01086483-v {girar}  
 01046072-v {desplazarse,ir,moveuse} 01049627-v {viajar} 01114042-v {ir,montar}  
 01046072-v {desplazarse,ir,moveuse} 01084973-v {andar,caminar}  
 01046072-v {desplazarse,ir,moveuse} 01122509-v {descender}  
 01046072-v {desplazarse,ir,moveuse} 01172741-v {pasar}  
 01054314-v {ir,marchar,partir,salir} 01147140-v {salir}  
 01055491-v {desplazar,mover} 00859635-v {poner}  
 01055491-v {desplazar,mover} 00869132-v {impeler,impulsar} 00867132-v  
 {arrojar,lanzar}  
 01055491-v {desplazar,mover} 00897572-v {separar} 00894185-v {cortar}  
 01212141-v {causar,hacer\_percibir} 01241976-v {hacer,sonar}  
 01256853-v {tener} 01256282-v {guardar,mantener,retener} 01268422-v  
 {conservar,preservar,resguardar} 00563886-v {consignar,hacer\_constar,registrar}

## Appendix IX FUE: working out the ‘communication’ semantic field

In this phase FUE has concentrated work in importing relations from WN1.5 - namely hypernymy, antonymy and meronymies for nouns, hypernymy and causation for verbs. This causes that most criticisms usually made to WN1.5 will apply as well to the resulting Spanish Database - i.e. too fine-grained unrelated senses for similar meanings, verb-to-noun unrelatedness, etc. Nevertheless, further use of other semantic relations defined within EWN can be used to diminish such effects, being the drawback that doing this seems to be highly cost-effective since either manual linking or manual revision of new links is probably going to be needed case by case.

In order to explore and show how using other relations the organization of the network can be improved a semantic field involving communication and related concepts (which is a semantic area specially confusing and not well solved in WN1.5) has been worked out. Following FUE’s approach no conflation of synsets-concepts has been made, that is, separate synsets as imported from WN1.5 remain as separate concepts in the Spanish Database: in the cases that close similitude of meaning would suggest two meanings could be conflated a near\_synonym link has been set between them.

The semantic field which has been worked out involves the following subtaxonomies (meaning of each concept is coarsely glossed in English; arrows indicate hyponymy steps):

### VERBS:

00333362-v saber, conocer (know, possess information)

00557762-v dirigirse, hablar (adress to an audience etc.)

00529407-v expresar, decir\_verbalmente (express)  
 ==> 00569629 afirmar, decir, declarar (say, declare...)

(act ==> interact ==>)

00416793-v comunicarse (transmit thoughts, feelings...)  
 ==> 00530290-v decir, hablar (verbalize, express in speech)  
 ==> 00467082-v comunicar, informar (inform, let know...)  
     ==> 00528672-v explicar (explain...)  
 ==> 00522332-v transmitir (make know, pass on -information)  
     ==> 00531321-v expresar, mostrar (show, e.g. disagreement)  
     ==> 00537777-v querer\_decir, significar (mean, intend to express...)  
 ==> 00542186-v hablar (speak, exchange thoughts verbally)  
 ==> 00559904-v escribir (write)

(create ==> re-create ==>)

00966090-v representar (represent, create an image of)

## NOUNS:

(abstraction ==&gt; relation ==&gt; social\_relation ==&gt;)

00018599-n comunicación (communication)

==&gt; 04187642-n lenguaje\_escrito (written language or communication)

==&gt; 04195435-n escritura (reading matter in general incl. autograph etc.)

==&gt; 04197046-n material\_escrito, lectura (e.g. books, magazines)

==&gt; 04211005.n texto (text)

==&gt; 04425761-n señal, signo (sign)

==&gt; 04434881-n símbolo (symbol)

==&gt; 04443305-n gráfico, símbolo\_gráfico (graphyc symbols)

==&gt; 04444555-n carácter, grafía (written symbols)

==&gt; 04451043-n carácter, letra (alphabetic letters)

==&gt; 04313427-n contenido, mensaje (message, content)

==&gt; 04337839-n información (information)

==&gt; 04155501-n lenguaje, comunicación\_lingüística (linguistic communication)

(artifact ==&gt; instrument ==&gt; means ==&gt;)

04140264-n medio\_de\_comunicación (medium)

==&gt; 04145615-n medio\_de\_comunicación\_escrita (print media)

==&gt; 04308479-n publicación (publication: books, reissues...)

==&gt; 04222100-n libro (book)

(artifact ==&gt; instrument ==&gt; means ==&gt; creation)

02354709-n representación (representation, visual rendering of sthing. or smone.)

(event ==&gt; happening ==&gt;)

044252761-n son, sonido (sound)

(act, human action ==&gt;)

04138929-n comunicación (communication, the act)

(entity ==&gt; ... human ==&gt;)

05842570-n comunicante (communicator)

==&gt; 06438760-n autor, escritor (author,writer)

New relations set between concepts are detailed below. For ease of explanation only one-way relation is shown (reversals must be inferred by the reader).

The first thing to be done is relating near-synonym concepts probably unnecessarily repeated in WN1.5, specifically concepts involving either 'speech' or 'communication' which show very subtle differences between them. This is done by using the 'near\_synonym' relation which allows us to both (i) preserve the WN1.5 concept distinction required by our approach; and (ii) allows every concept to retain its own set of hyponyms:

(a) 00529407-n {expresar,decir\_verbalmente} near\_synonym 00530290-n {decir,hablar}

This way hyponyms of ‘00529407-n {expresar,decir\_verbalmente}’ get indirectly related to ‘transmit thoughts, feelings...’

(b) 00530290-n {decir,hablar} near\_synonym 00542186-n {hablar}

Thus relating the more communicative-like ‘exchange thoughts’ and the ‘verbalize’ sense of ‘speak’.

(c) 00467082-v {comunicar,informar} near\_synonym 00522332-v {transmitir}

This link stands for the practical identity of meaning between ‘inform, let know’ and ‘make know’. The difference between both concepts in WN1.5 seems to rely in a stress in the volition of the agent in the case of 00522332-v {transmitir}.

(d) 00018599-n {comunicación} near\_synonym 04138929 {comunicación}

Thus relating the ‘social\_relation’ and the ‘human\_action’ senses of ‘communication’. Moreover, crossing POS...

(e) 04138929 {comunicación} xpos\_near\_synonym 00416793 {comunicarse}

... a verb and an action noun which refer to the same event.

Moreover, co-hyponymy between ‘written language’ and ‘linguistic communication’ is re-defined: it seems better that written language is a way of linguistically communicating, so:

(f) 04187642 {lenguaje\_escrito} has\_hyperonym 04155501 {lenguaje, comunicación\_lingüística}

(the link ‘04187642 {lenguaje\_escrito}’ has\_hyperonym ‘00018599-n {comunicación}’ is removed; so hyponyms of ‘written language’ will be now hyponyms of ‘linguistic communication’... and also, via ‘linguistic communication’, still hyponyms of ‘communication’)

Furthermore, unrelatedness of ‘00557762-v dirigirse, hablar (adress to an audience etc.)’ and ‘00530290-v decir, hablar (verbalize, express in speech)’ is solved by stating hyponymy between one and another:

(g) 00557762-v {dirigirse, hablar} has\_hyperonym 00530290-v {decir, hablar}

Some relations of causation non present in WN1.5 stand between ‘communicate’ events and related actions. The following have been set:

- (h) 00966090-v {representar} causes(non-factive) 00416793-v {comunicarse}  
 (i) 00530290-v {decir, hablar} causes(factive) 00467082-v {comunicar, informar}  
 (j) 00467082-v {comunicar, informar} causes(non-factive) 00333362-v {saber, conocer}

Roughly, relations above express the idea that express in speech causes (always) let know; and that letting know, in turn, usually (but not always) causes to know.

Some involvement relations also stand between communicative events (expressed either by nouns or verbs) and participants. The following have been set:

- (k) 00966090-v {representar} involved 02354709-n {representación}

(the result of creating an image is the image thus created; this resultative relation can not be explicitly encoded using actual EWN relations, so we use the underspecified relation of involvement)

- (l) 06438760-n {autor, escritor} role\_agent 00559904-v {escribir}  
 (m) 00559904-v {escribir} involved 04187642-n {lenguaje\_escrito}

(a writer is the agent of the action of writing; the result of writing is writing language).

- (n) 05842570-n {comunicante}role\_agent 00416793-v {comunicarse}

(the action of transmitting thoughts, feelings etc. requires a communicator)

- (o) 04138929-n {comunicación}involved 04313427-n {contenido, mensaje}

(the action of communication involves the existence of a message or content of the communication)

- (p) 04140264-n {medio\_de\_comunicación} role\_instrument 00416793-v {comunicarse}

(this way communicative media are related to communication: they are instruments for achieving event -but not necessarily the event is always achieved)

A couple of part-whole relations seem to stand within the semantic field:

- (q) 044252761-n {son, sonido} is\_subevent\_of (reversed) 00530290-v {decir, hablar}

(a series of 'sound' events are parts of the whole event of 'expressing in speech')

Last, a nominal part-whole relation imported from WN1.5 has been re-defined:

- (r) 04211005-n {texto} has\_holo\_part 04145615-n {medio\_de\_comunicación\_escrita}

thus expressing that newspapers etc. have text in them. The original WN1.5 equivalent relation was between 'text' and 'publication' (an hyponym of 'print media' which stands for books and so); this relation too-low mapping has been removed since not only books but also the rest of print media (newspapers, magazines) have text as part.

Both new and re-defined links described above build up a more explicative organization of this area of the lexicon than that existing in WN1.5, thus allowing for 'surfing' between related concepts. Unfortunately most of such relationships seem hard to acquire by (semi)automatic means -sources are poor and full of ambiguity. Nevertheless, having time and human resources enough a fine conceptual network could be built using the theoretical framework defined within EWN.