View metadata, citation and similar papers at core.ac.uk
brought to you by CORE
provided by DSpace at VU

# The Dutch Wordnet

Version 2, Final
July 12, 1999

Piek Vossen , Laura Bloksma, Paul Boersma
*University of Amsterdam*

# Table of Contents

# List of Tables and Figures

# 1. Introduction

This document describes the Dutch wordnet developed by the University of Amsterdam in the EuroWordNet project (LE2-4003). The EuroWordNet database has been built from available existing resources and databases with semantic information developed in various projects. In our case, we used a database for Dutch provided by Van Dale Publishers BV. This database already has semantic relations between word senses and provides an excellent starting point for developing a Dutch wordnet. The document is divided into the following sections:

- section 2 describes the main resources that we used;
- section 3 discusses the methodology;
- section 4 describes the structure and content of the Dutch wordnet;
- section 5 describes the content of the CD for the Dutch wordnet;

In the Appendix, examples are given of entries in the text files for the different parts-of-speech (nouns, verbs, adjectives/adverbs). For an explanation of the structure we refer to the general EuroWordNet documentation on the general CD or downloadable from the www-site http://www.hum.uva.nl/~ewn. The general document describes the design of the multilingual database, the language internal relations and the equivalence relations. It also includes a specification of the overall approach for building the wordnets and the top-ontology that is used as a common framework. In the following, we assume that the reader is familiar with this general EuroWordNet document.

The Dutch wordnet consists of 34,455 nominal synsets, 9,040 verbal synsets and 520 adjectival/adverbial synsets, where the latter are only added if they are relevant for the meaning of the encoded nouns and verbs. Each synset consists of synonyms, a set of language-internal relations and, mostly, an equivalence relation to the closest WordNet1.5 synset. The synsets do not have glosses or definitions but a classification in terms of the EuroWordNet top-ontology.

The Dutch wordnet is property of the University of Amsterdam and Van Dale Publishers BV and can be licensed via ELDA/ELRA (http://www.icp.grenet.fr/ELRA/home.html). For any further information on the Dutch wordnet please get in touch with:

Dr. Piek Vossen
Co-ordinator of EuroWordNet
Faculteit der Geesteswetenschappen
Universiteit van Amsterdam
Spuistraat 134
Amsterdam, 1012 VB
NL
tel: +31 20 525 4669
fax: +31 20 525 4429
e-mail: piek.vossen@hum.uva.nl

## 2. The resources used for building the Dutch wordnet

We used the following resources for the Dutch wordnet:

- Vlis: the content of a lexical database provided by Van Dale;
- Vde: the Van Dale Dutch-English dictionary (Martin and Tops 1986);
- Ved: the Van Dale English-Dutch dictionary (Martin and Tops 1989);
- WordNet 1.5 (Fellbaum 1998);
- Celex Dutch and English lexicons with basic morpho-syntactic information and corpus frequency information on word forms and lemmas;
- a list of new Dutch spellings (according to the spelling convention of 1997) extracted from a Dutch-Russian dictionary build at the University of Amsterdam;

We assume that the reader is familiar with WordNet1.5. The main resources used are Vlis, the Vde and the Ved. Some background information will be given for these resources in the next sections. The Celex lexicons and the list of new spellings will not be further discussed. All the above resources were loaded in a special lexical database system developed at the University of Amsterdam (the Amsterdam Lexicon System or ALS). A more complete description of the functionality of ALS and the lexical data structures for the above resources is given in the ALS manual (downloadable from the EuroWordNet WWW-site).

### *2.1. The Vlis database*

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

*Table 1: Number of entries and senses in the Vlis database*

| | *Nouns* | *Verbs* |
|---|---|---|
| **Entries** | 63962 | 8822 |
| **Senses** | 74678 | 14268 |

The coverage is contemporary Dutch and it includes the common and general vocabulary as most common monolingual dictionaries. A fundamental difference with traditional dictionaries is however that particular semantic and morphological relations are explicitly coded for nouns, verbs and adjectives at a sense level. The next table displays the various relations distributed over nouns and verbs (the adjectives are left out here).

*Table 2: Semantic relations in the Vlis database*

| | *Noun senses* | *Verb senses* |
|---|---|---|
| **ABBREVIATION** | 63 | 0 |
| **ANTONYM** | 181 | 148 |
| **ASSOCIATIVE** | 1763 | 557 |
| **CAUSATIVE** | 0 | 5 |
| **HYPERONYM** | 49815 | 8231 |
| **INCHOATIVE** | 0 | 11 |
| **PARTITIVE** | 576 | 0 |
| **REFERENCE** | 715 | 93 |
| **SYNONYM** | 9901 | 5049 |
| **WOMAN** | 95 | 0 |
| **PREFERENCE** | 2361 | 108 |
| **FORM_VARIANT** | 4 | 0 |
| **SYMBOL** | 1 | 0 |

Important semantic relations in Vlis are hyponomy, synonymy, antonymy, partitive and associative. Hyponymy is a 'kind of' relation from a hyponym (e.g. “auto” (car)) to its class or hyperonym (e.g. “motorrijtuig” (motor vehicle)) and vice versa. The difference in meaning between co-hyponyms that share the same hyperonym is formulated in a field called 'differentiae'. Whereas WN1.5 occasionally includes multiple hyperonyms this does not occur in the Vlis database (but has been added to Vlis for the Dutch wordnet).

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 '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 only assigned to nouns. The relation is used as the meronymy relation in WN1.5, but in Vlis it is not differentiated into subtypes. Associative is always an additional relation between two senses that are closely related and is allowed between all part-of-speech categories. The idea was to assign the relation if other relations 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 causative 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 variants to the synonymy or hyponymy relation.

The database does not form a *closed* wordnet in which all relations are unified in a single tree or a small set of tops. In fact, the Vlis database contains 1429 tops for nouns and 298 tops for verbs. The reason for making a sense a top in the hierarchy is often not well-motivated (the database is still in development). In many cases a sense is an end point because the meaning was too complex (e.g. higher-order nouns referring to states, conditions, events) or information was of an encyclopaedic nature (names of places, people, etc.).

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 for each word form (per part-of-speech) extracted from the Celex database.

### *2.2. The Van Dale Dutch-English and English-Dutch Dictionary*

The Van Dale bilingual dictionaries are developed for native speakers of Dutch. This means that the resources contain only very limited information on the Dutch words and much more information on the foreign-language target words. The data for the Dutch-English and English-Dutch dictionaries (Martin and Tops 1986 and 1989) is stored in the form of separate fields with field-names and values. Some values are restricted to codes, others contain free text. The entry-structure is homograph-based but homographs are distinguished only when the part-of-speech differs and/or the pronunciation. Sub-homographs are used when senses differ in major grammatical properties such as valency, countability, predicate/attributive usage.

*Table 3: Number of entries, senses and translations in the Van Dale Dutch-English & English-Dutch dictionaries*

| | *Dutch-English* | *English-Dutch* |
|---|---|---|
| **Entries** | 90,925 | 89,428 |
| **Senses** | 127,024 | 156,838 |
| **Main Translations** | 145,511 | 152,318 |
| **Secondary Translations** | 104,181 | 162,752 |

In addition to some grammatical information on the words and the translations, the dictionary contains a large amount of semantic information restricting the senses and the translations. In the case of the Dutch-English dictionary, we find for example the following additional information:

- [Sense-indicators] (53368 tokens) to specify the Dutch senses or polysemous entries. These contain bits and pieces from original definitions (often a genus word);
- [Biological gender marker] for English translations. This is necessary to differentiate translations when the source and target language have different words for male or female species: 286 translations are labelled as male, 407 translations as female;
- [Usage labels for domain, style and register] Applies to both Dutch senses and their English translations;
- [Dialect labels] for Dutch senses and their English translations;
- [Context markers] (23723 tokens, 16482 types). These are semantic constraints differentiating the context of multiple translations, and to limit the scope of translations having a narrower context than the Dutch source sense;

The usage labels and the domain labels are mostly stored in the same field. Differentiation has to be done by some parsing. The usage labels form a limited closed set of abbreviations and codes, the domain labels are free text. For the main-translations, about 400 different types of usage labels.

The translations can be single words, words combined with labels, co-ordination of translations and phrases. Phrasal translation may indicate a lexical gap in English or point to a multiword expression in the target language. Co-ordinations have been marked in the resource by "//" (for alternative words) or "/" (surrounding alternative phrases). This information has been used to

split them in separate translation fields for a sense, e.g.:

gin//genever bottle
=>gin bottle; genever bottle
(administration of) the /last sacraments/extreme union/
=> administration of the extreme union; administration of the last sacraments; the last sacraments; the extreme union
(adult) literacy project//campaign
=> adult literacy project; adult literacy project; literacy project; literacy campaign

The result is a rather basic resource with sense-indicators for Dutch and English senses and one or more translations per sense, where each translation field is a phrase or a single word.

## 3. Methodology

In the general EuroWordNet documentation, you can find information on the overall methodology and the design of the EuroWordNet database. To explain the strategy for the Dutch wordnet, we will summarise the main points here. As discussed in the general EuroWordNet document, the design of the database is such that the wordnets can be developed and maintained relatively independently. The main motivation for this is:

1. the wordnet builders have different resources and database and tools for building the wordnets;
2. the lexicalization of words differs from language to language, resulting in different wordnet structures for each language;

Globally, there have been 2 main approaches for building the wordnets in EuroWordNet:

1. expand approach: WordNet1.5 synsets are translated to another language and the WordNet1.5 relations are taken over as a start;
2. the merge approach: a separate wordnet is build up for a language with a unique structure and this wordnet is then mapped to WordNet1.5 by generating equivalence relations;

The expand approach results in structures that are close to WordNet1.5 but may also be biased by it. The merge approach is more difficult and may result in very different structures.

For the Dutch wordnet, we followed the merge approach, mainly because we already had a structured database with relations between word senses and we are particularly interested in differences of the lexicalization patterns in Dutch and other languages. A similar approach has been followed for Italian, whereas the Spanish wordnet has been built according to the expand method.

A drawback of the flexible design and construction of the database is that specific measured had to be taken to guarantee a minimal compatibility in coverage and interpretation across the wordnets. This has been achieved by adopting general top-down approach where the wordnets have been built starting with a set of about 1300 common Base Concepts (see the general EuroWordNet documentation for a further explanation). These Base Concepts, which are represented as Inter-Lingual-Index records (mostly WordNet1.5 synsets), play an important role in two or more wordnets. Importance is measured in terms of number of relations attached to

these concepts and the position in the hierarchy. They are generic concepts that express major generalizations over other word meanings. In this respect, they are more general than Basic Level concepts as defined by Rosch (1977). The latter typically occur at a middle level of specificity.

Each wordnet builder took this set of Base Concepts as a starting point and tried to represent the concepts in their local wordnet. The next step was to build a core wordnet around these Base Concepts, possibly extended with concepts that are important in the local wordnet but have not been selected as common Base Concepts.

The core wordnets include all relevant relations to the local equivalents of the Base Concepts, minimally consisting of:

- hyperonyms;
- 1st level of hyponyms below the Base Concepts;
- equivalence relations;

The core wordnets have mainly been built manually and there has been extensive discussion on the nature of the relations. Furthermore, the EuroWordNet top-ontology has been applied to all the Base Concepts. The EuroWordNet top-ontology provided all the wordnet builders with a common semantic framework and made it possible to monitor the progress by clustering synsets over the top-concepts (e.g. *Animal*, *Vehicle*, etc.). For the rest, each builder was free to extend the core wordnets to the final wordnets, either using the above expand or merge approach. Extension of the core wordnets has been done (semi-)automatically.

In general, all the wordnets are based on the lexicalizations of nouns and verbs in each language. Multiword expressions are allowed but have not been the main focus. The minimal size of the wordnets should be about 35K synsets and 50K word meanings. The vocabularies should include:

- all the generic word meanings that are needed to properly link and classify more specific meanings in a language;
- all most frequent words in a language, as measured in a general language corpus;

The first criterion is ensured by the above top-down approach. The second criterion has been tested by comparing the entries in the wordnets with Parole lexicons. In the Parole project, morpho-syntactic lexicons have been built for the 20,000 most frequent words in 12 European languages, based on comparable corpora of general language.

The Dutch wordnet has also been built according to this top-down approach, starting with the lexical semantic relations as they occur in the Vlis database. The main aim has been to develop a consistent and complete Dutch wordnet, given the funding and time available to us. We mainly used language-internal criteria, tests and definitions as they have been agreed upon in EuroWordNet, but we also frequently compared the Dutch structures with WordNet1.5 to choose the best solution. The second aim was to link this Dutch wordnet to WordNet1.5.

The development of the Dutch wordnet can then roughly be described by the following steps:

1. conversion of the Vlis database to the EuroWordNet structure and the addition of the Dutch database with Celex corpus frequency information;
2. automatic generation of equivalence relations via the bi-lingual dictionaries;
3. the development of the Dutch core wordnet around the Dutch equivalences of the common Base Concepts and other Dutch concepts that are important;
4. extension of the Dutch core wordnet to the full wordnet;

Step 1 and 2 resulted in a large database loaded in the Amsterdam Lexicon System (ALS) that contained a basic EuroWordNet (EWN) structure with language-internal relations between word senses and an automatic equivalence mapping to WordNet1.5. This database has been used to make the first selections of Base Concepts and to build the core wordnet of about 10,000 concepts (step 3). The core wordnet has been compared with the other wordnets, with corpus frequency information and by measuring the clustering with respect to the top-ontology and the associated Base Concepts. On the basis of this comparison we have improved the quality of the data and extended the selection to full coverage. Since a new Dutch spelling has been introduced in 1997, we have also added new spelling variants to the selection. The four main steps will be further described in the next subsections

### *3.1. Conversion of the Vlis and Celex to the Dutch EuroWordNet structure*

Because of the partial compatibility of the relations in Vlis and the relations defined in EuroWordNet, we decided to convert the Vlis-relations to the corresponding EuroWordNet relations and use this database as a starting point. Table 4 indicates how the Vlis relations have been converted.

*Table 4: Conversion of the Vlis relations*

| | *Noun senses* | *Verb senses* | *Conversion to Dutch wordnet* |
|---|---|---|---|
| **ABBREVIATION** | 63 | 0 | Synset member |
| **ANTONYM** | 181 | 148 | NEAR_ANTONYM |
| **ASSOCIATIVE** | 1763 | 557 | -- |
| **CAUSATIVE** | 0 | 5 | CAUSES |
| **HYPERONYM** | 49815 | 8231 | HAS_HYPERONYM |
| **INCHOATIVE** | 0 | 11 | CAUSES |
| **PARTITIVE** | 576 | 0 | HAS_MERONYM |
| **REFERENCE** | 715 | 93 | -- |
| **SYNONYM** | 19901 | 5049 | Synset member |
| **WOMAN/MAN** | 95 | 0 | Synset member |
| **PREFERENCE** | 2361 | 108 | Synset member |
| **FORM_VARIANT** | 4 | 0 | Synset member |
| **VERB** | 0 | 474 | HAS_HYPERONYM |

The conversion resulted in a first Dutch wordnet. Next we loaded the Celex Dutch lemma lexicon with frequency information in ALS and we have added frequency information to the matching Vlis entries with the same part of speech. Finally, we created slots for the specification of equivalence relations to WordNet1.5 synsets and for specifying EuroWordNet top-concepts that apply.

Figure-1 shows the resulting structure of the database in ALS for the first sense of "werktuig" (tool). First, the part-of-speech and gender is given (NOUN_HET), followed by the Celex frequency (f = 328). and the second line contains a definition ("stuk gereedschap" *piece of equipment*). The Upward and Downward relations are the relations given in Vlis. After that the relations according to EuroWordNet are listed. First, the other synset members, next, the language-internal relations (EWN relations), the equivalence relations (EWN inters) and the top-concepts (EWN top concepts). The shown information can be customized, e.g. here the HYPONYMs relations are not expanded to reduce the amount of information in the scrollable window.

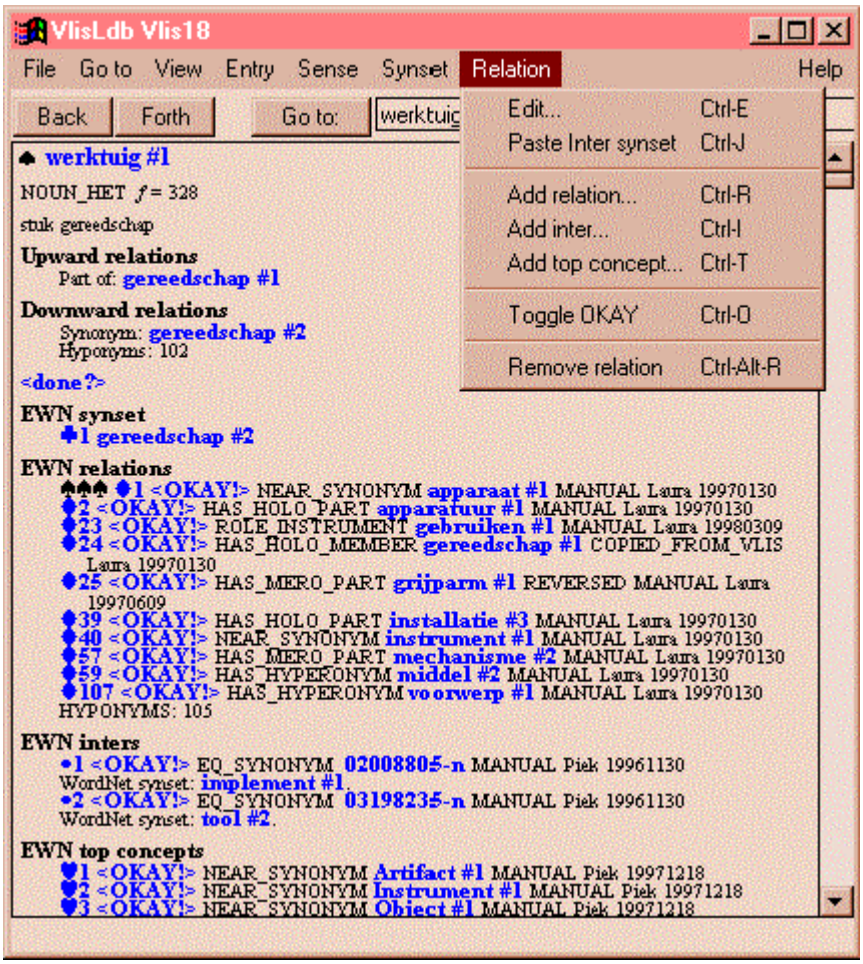


*Figure 1: Window for the VlisEntry "werktuig" (tool) in the Amsterdam Lexicon System (ALS)*

As the opened menu shows, there are different possibilities to modify the relations. For example, copied relations can be checked manually, and if agreed, marked as being <OKAY>, and relations can also be removed, added or edited. Added relations are automatically reversed from the target sense back to the sense where the relation is added. The Dutch wordnet thus consists of relations which can have to the following status, as is also shown in Figure-1:

1. relations copied from Vlis and not manually checked (22% of the relations in the Dutch wordnet);
2. relations copied from Vlis and manually confirmed as OKAY (47% of the relations);
3. relations added manually (24% of the relations);
4. relations added manually and confirmed as OKAY (7% of the relations);

Obviously, Vlis relations also have been removed. This is not shown in the above distribution. For a further explanation of the ALS functionality, see the ALS manual (Vossen, Boersma and Bloksma 1999).

## *3.2. The extraction of equivalence relations to WordNet1.5.*

The extraction of the equivalence links between the Dutch wordnet and WordNet1.5 is partly done by hand and partly using automatic techniques. The manual coding is carried out for the most important concepts in the database (see below) and for those concepts that have been poorly matched by the automatic techniques. The manual encoding of the most important concepts had to ensure that the cores of the wordnets are well-matched. In total, 14,749 equivalence relations (30% of all equivalence relations) have been created by hand or confirmed manually (8,596 manual equivalence relations for nominal synsets and 6,153 manual equivalence relations for verbal synsets). These figures include the manual encoding of poor translations and the manual confirmation of automatically generated translations, which will be further discussed next.

### *3.2.1. Extracting translations from bilingual dictionaries.*

Most synsets have been translated by mapping the Van Dale database with the bilingual Dutch-English dictionary and mapping the translations to WordNet1.5. This implies that the part-of-speech of the orthographic form in Vlis and in the bilingual dictionary match, as well as the part-of-speech of the translations in WordNet1.5. Note that such a mapping is carried out for synsets. It is therefore sufficient if one member of a synset can be mapped and there may be multiple mappings originating from different synset members.

A proportion of the original Dutch database did not receive a translation, either because the entry was missing in the bilingual dictionary, or the translations could not be found in WordNet1.5.

*Table 5: Synsets without translation in the Dutch database*

| | *Number of Synsets in Vlis* | *Number of Synsets without translation to WordNet1.5* | *%* |
|---|---|---|---|
| **nouns** | 52359 | 23398 | 44,69% |
| **verbs** | 9125 | 1060 | 11,62% |

Table 5 shows that the result for verbs (11% not translsated) are much better than for nouns (44% not translated). This is due to the fact that the nominal part contains more specialized vocabulary. To improve this matching we have applied two additional techniques.

According to the introduction to the Dutch-English dictionary, many English words that are directly taken over in Dutch without change of meaning and pronunciation, have been omitted (to save space, assuming that Dutch speakers are familiar with the word). It therefore makes sense to directly match the non-translated Dutch entries to the WordNet1.5 entries. The results are given in the next table:

*Table 6: Directly matching entries between the Dutch database and WordNet1.5.*

| | *Entries in Vlis* | *Entries in WordNet1.5* | *Intersecting Entries* |
|---|---|---|---|
| **nouns** | 63962 | 88200 | 3981 |
| **verbs** | 8822 | 14734 | 9 |

Inspection of this list showed that all the 9 verbal matches were wrong but the nominal intersection contained many good matches. By intersecting the direct nominal matches with the synsets without translation the nominal matching has been improved as follows:

*Table 7: New nominal translations generated by direct matches with WordNet1.5*

| | *Synsets without translation* | *Senses without translations* | *Entries without translations* | *Overlap with WordNet1.5 noun entries* | *Matched synsets* | *Remaining unmatched synsets* |
|---|---|---|---|---|---|---|
| **nouns in the Van Dale database** | 23398 | 27894 | 27053 | 841 | 726 | 22672 |

A second improvement consisted of reversing the English-Dutch dictionary (Ved) into a Dutch-English dictionary, assuming that the set of Dutch translations is different from the set of Dutch entries. A third improvement consisted of varying the use of hyphens and spaces in the translations. In many cases, hyphens and spaces are used inconsistently, e.g. *animal park*, *animal-park* and *animalpark*. By replacing and removing spaces and hyphens in the translations we could further translate another 338 synsets . This resulted in the following improvements:

*Table 8: Translations by reversed English-Dutch dictionary and replacing spaces and hyphens.*

| | *Total Number of Synsets* | *Synsets without translation* | *Synsets matched by reversed English-Dutch dictionary* | *Synsets matched by replacing Spaces and Hyphens* | *Remaining unmatched synsets* | *% of total* |
|---|---|---|---|---|---|---|
| **nouns** | 52359 | 22672 | 2161 | 338 | 20180 | 38.54% |
| **verb** | 9125 | 1060 | 183 | 7 | 869 | 9.52% |

Since not all nouns from the Van Dale database are selected for the Dutch wordnet, the result for the wordnet is better:

*Table 9: Synsets without translation in the Dutch wordnet*

| | *Dutch WordNet* | *No ILI match* | *% of Dutch Wordnet* |
|---|---|---|---|
| **nouns** | 34455 | 6070 | 17.6% |
| **verbs** | 9040 | 1133 | 12.5% |

For the Dutch wordnet, 82% of the selected noun synsets and 87% of the verb synsets has a match to a WordNet1.5 synset. The fact that this result is better than for the complete database has to do with the way synsets have been selected. First of all, we selected synsets with manual and reliable translations and, secondly, we excluded more specific levels that are more likely not translated (and probably cannot be translated). Note that this is the final matching result, which includes the manual addition of complex equivalence relations and the removal of wrongly-generated matches. In the case of the verbs, the results are lower. This is because many wrong automatic translations have been removed or neglected (see below).

### *3.2.2. Weighting the translation candidates*

Once a matching entry has been found in WordNet1.5, all the senses of the entry are proposed as possible translations. If there is only one synset translation, the procedure stops, and we assume that this translation is correct. If there are multiple translations, they are weighted using several heuristics:

1. prefer translations of Dutch senses in the bilingual dictionary that match the information given for the Dutch sense in Vlis;
2. conceptual distance measurement: prefer translations that are close in the WordNet1.5 hierarchy;
3. overlap with Dutch synsets when target WordNet1.5 synsets are translated back to Dutch;
4. overlap in Top-Concepts inherited according to the Dutch hierarchy for the Dutch sense and according to the WordNet1.5 hierarchy for the target translations;

The first heuristics makes use of the fact that the bilingual dictionaries often give some further information on the Dutch sense when it is polysemous. In many cases, this information consists of some key words from the definition in the monolingual dictionary, but sometimes usage labels or additional information are given. By matching the overlap in characters in the definitions relative to the length of the definition, an initial weighting is calculated. Furthermore, there are some more specific morpho-syntactic features that can be used to select the most appropriate sense in the bilingual dictionary, e.g. valency for verbs, gender for nouns. If this information is present for both the Dutch synset in the monolingual resource and the bilingual entry, then only translations of senses with matching information are considered. For the rest all possible target synsets of the translations are listed (in some cases hundreds of synsets). The output of the matching of Vlis to the bilingual resource is used to for further weighting.

The second heuristics makes uses of the notion of conceptual distance as defined by Agirre and Rigau (1996). Conceptual distance is calculated by counting the steps to their closest shared node in the network, taking into account the level of the hierarchy and the density of nodes relative to the average density. There are two situations for which the conceptual distance is calculated:

1. the distance between senses of multiple alternative translations of a single entry in the bilingual dictionary.
2. the distance between each possible translation and the translations of hyponyms and hyperonyms of the Dutch word

The first situation occurs when, for example, the Dutch word *orgel* has two translations, *organ* and *keyboard*, for the same sense. Since the polysemy of these translations is often not parallel, it is possible to favour the sense of *organ* and *keyboard* that have the shortest distance. The second situation is illustrated in Figure-2. Here we see that *orgel* in Dutch is translated as *organ*, which can either be a *musical instrument* or a *body part*. Since the hyperonym and a hyponym of *orgel* in the Dutch wordnet have already been translated it is possible to measure the distance of the two senses of *organ* to the translations of the hyperonym and hyponym:

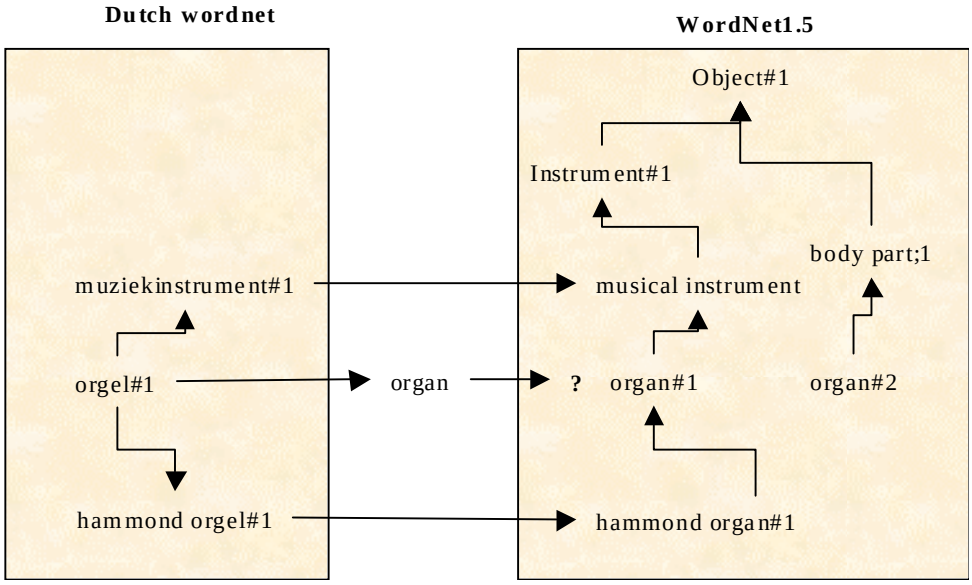


*Figure 2: Selecting translations to WordNet1.5 by distance to the translated context in the Dutch wordnet.*

The distance measuring of the translations to the context in the Dutch wordnet, leads to a ranking of all the senses of a translation. The heuristics for automatically deriving equivalence relations are implemented in such a way that bad matches are removed if the best match is above a certain threshold. If not, all matches are maintained. A large number of matches with a low score therefore indicates that the system had poor evidence for differentiating the matching. By searching the database for synsets which have an extremely high number of automatically-derived translations (with a relatively low score) we can isolate dubious cases. This is shown for the next example "inlassen" (to weld something in between something else), where none of the suggested translations is correct (in fact, the correct translation probably does not exist):

Dutch synset: inlassen \# 2

| | | |
|---|---|---|
| 2.07 | 00604079-v | bring out#3; introduce#6 |
| 1.77 | 00121079-v | alter#4; falsify#1; interpolate#1 |
| 1.7 | 00579406-v | insert#5; slip in#1; sneak in#1; stick in#2 |
| 1.65 | 00437968-v | barge in#1; break in#4; butt in#1; chime in#1; cut in#4; put in#2 |
| 1.65 | 00514811-v | come in#2; inject#3; interject#1; interpose#1; put in#3; throw in#1 |
| 1.35 | 00361286-v | extrapolate#2; interpolate#2 |
| 1.30 | 00818159-v | enter#3; infix#3; insert#7; introduce#7 |
| 1.27 | 01417019-v | admit#4; allow in#1; let in#2 |
| 1.13 | 00507610-v | introduce#5; preface#2; premise#3 |
| 1.00 | 00799930-v | insert#6; tuck#3 |
| 0.946 | 00927659-v | introduce#8 |
| 0.946 | 00939471-v | innovate#1; introduce#9 |
| 0.914 | 00507320-v | acquaint#2; introduce#4; present#6 |
| 0.898 | 00397690-v | introduce#3 |
| 0.668 | 00210341-v | inaugurate#1; introduce#2; usher in#1 |
| 0.668 | 01189328-v | bring in#2; introduce#10 |
| 0.668 | 01297479-v | hive away#1; lay in#1; put in#5; salt away#1; stack away#1; stash away#1; store#7 |
| 0.668 | 01532350-v | put in#6 |
| 0.544 | 01386819-v | admit#3; include#3; let in#1; let participate#1 |
| 0.364 | 00113224-v | insert#4; introduce#1; put in#1; stick in#1 |
| 0.243 | 00605466-v | put in#4; submit#5 |

The number-codes, such as 00605466-v and 00113224-v, are file-offset positions that uniquely identify a synset in WordNet1.5.

Based on the above observations, we have manually translated all verbal synsets with more than 20 translations and all nominal synsets with more than 30 translations. Next we looked at:

- polysemous words with many meanings and many translations
- synsets with many relations and many translations

It often appeared that polysemous words with a badly translated sense also had poor translations for the other senses. We then manually translated all the senses of such a polysemous word. In addition, we have looked at words with many relations and many translations. All verbs with more than 2 relations and more than 10 translations have been manually translated as well. The same holds for nouns with more than 10 relations and more than 10 translations. About 3,000 synsets with low quality matches have thus been translated by hand.

The next step was to rerun the tree-matching algorithm. The matching algorithm tries to weight candidate translations by calculating the distance in the WordNet1.5 hierarchy of each translation to the translations of the Dutch hyponyms and hyperonyms. Since many translations have been improved manually, we expected an improvement of the tree-matching effect for synsets related to these concepts (by hyponymy or hyperonymy) as well. Table 11 below shows the results.

The third heuristics makes use of the possibility to access the Dutch-English and English-Dutch dictionary vice versa and to prefer translations that co-occur and show overlap with the original Dutch synset. If several synset members in WordNet1.5 have the same Dutch word as a translation in the English-Dutch dictionary or they have several Dutch translations in the same Dutch synset, then this can be seen as additional evidence for the correctness of a translation. Roughly the algorithm is:

1. take the possible candidate translations generated from the Dutch-English resource to WordNet1.5;
2. look up the target variants in the English-Dutch resource;
3. increase the match:
   3.1. each time an English variant has a variant of the Dutch source synset as its translation;
   3.2. if multiple variants of the Dutch source synset are given as the translation for a single English sense;

The next example illustrates this for the Dutch synset "lakken" (coat with lacquer):[1]

```
Dutch Variants: lakken;
WordNet Match: 00779724-v
        WordNet Variant: affix a seal to
        WordNet Variant Translation:
        WordNet Variant: seal
        WordNet Variant Translation:  op robben/zeehondenvangst gaan/zijn;
        Overlap = 0
WordNet Match: 00726098-v
        WordNet Variant: coat with lacquer
        WordNet Variant Translation:
        WordNet Variant: lacquer
        WordNet Variant Translation:  lakken;
        WordNet Variant Translation:  vernissen;
        Overlap = 1
        From: 26.6  To: 39.8
```

The verb "lakken" has a translation candidate synset 00779724-v. The first variant of this synset is "affix a seal to". This cannot be found as an entry in the English Dutch dictionary. The second variant "seal" can be found. However, none of the translations of "seal" contains the original Dutch word "lakken" (the overlap between the translations and the original synset members is 0). The next WordNet Match is the synset 00726098-v. The first synset member "coat with lacquer" cannot be found but the second "lacquer" is found and has "lakken" as one of its translations. The matching for this synset will thus be increased (From: 26.6  To: 39.8). The second translation of “lacquer” is a near synonym “vernissen”. Note that if “vernissen” would be part of the Dutch synset “lakken”, this would result in an overlap of 2 but also cause an extra increment because both synset variants are given for a single sense of “lacquer” in the bilingual English-Dutch dictionary.

The fourth heuristics makes use of the fact that we have separately added the 63 Top Concepts (TCs) to the Base Concepts (the most important concepts) in the Dutch wordnet and WordNet1.5. The TCs represent fundamental semantic features, such as Natural, Artifact, Dynamic, Static, Physical, Mental that can be combined into complex feature combinations (see the general documentation for a further motivation and explanation of the top-ontology). By inheriting these TCs to more specific concepts via the hyponymy relations it is possible to measure the overlap in TCs between Dutch senses and their candidate translations. If a candidate translation has many overlapping TCs, it is a more likely candidate for translating. In the next example, the Dutch word "hart" (heart as an organ) inherits the top-concepts *Living* and *Part* from its hyperonyms (orgaan 1, deel 2, iets 1), which it shares only with sense 4 of the senses of "heart" in WordNet1.5:

[1] In the real situation there are many more matches and translations. Here, we have listed just two of them to illustrate the example.

hart 1
orgaan 1 (*Living Part*) deel 2 (*Part*) iets 1 LEAF
---
heart 1
playing card 1 card 1 (*Artifact Function Object*) paper 6 (*Artifact Solid*)
material 5 (*Substance)* matter 1 inanimate object 1 entity 1 LEAF
heart 2
disposition 2 (*Dynamic Experience Mental*) nature 1
trait 1 (*Property*) attribute 1 (*Property*) abstraction 1 LEAF
heart 3
bravery 1 spirit 1 character 1 trait 1 (*Property*) attribute 1 (*Property*)
abstraction 1 LEAF
heart 4
internal organ 1 organ 4 (*Living Part*) body part 1 (*Living Part*)
part 10 entity 1 LEAF

This heuristics is expected to be especially useful for discriminating translations of verbs because their semantics is less dependent on the hierarchical structure (which is relatively flat and shallow). A rich encoding with features for verbs with a poor hyponymic structure can still contain sufficient evidence for choosing translations. The effect of this matching obviously depends on the coverage of the features and the diversity of features. In EuroWordNet, we have limited ourselves to the 63 features from the EuroWordNet top ontology. To further improve the matching it is possible to add more discriminative features at crucial points of both hierarchies. To get a maximal coverage of inherited top-concepts, we ensured that all tops in the Dutch wordnet and in WordNet1.5 are classified according to the ontology, and that most tops in the Dutch wordnet are unified into a minimal number of trees. For WordNet1.5, we had to add top-concepts (TCs) to 389 verbal synsets and 2 nominal synsets, which are tops but have not previously been classified by the top-ontology. In total 2006 Dutch synsets (1170 nouns and 836 verbs) and 1410 WordNet1.5 synsets (793 nouns and 617 verbs) have been classified with one or more top-concept features. Furthermore, we have converted the lexicographer's file codes in WordNet1.5 to compatible EuroWordNet top-ontology codes, as is indicated in the next table. Since all synsets in WordNet1.5 have been assigned by these codes we thus get a very high coverage of the semantic features.

*Table 10: Conversion of WordNet1.5 Lexicographer's file codes to EuroWordNet top-concepts*

| Code | WordNet File Name | EuroWordNet Top Concepts | Code | WordNet File Name | EuroWordNet Top Concepts |
|---|---|---|---|---|---|
| **04** | noun.act | Agentive; | **29** | verb.body | Dynamic; Physical; |
| **05** | noun.animal | Animal; | **30** | verb.change | Dynamic; |
| **06** | noun.artifact | Artifact; | **31** | verb.cognition | Mental; Dynamic; |
| **07** | noun.attribute | Property; | **32** | verb.communication | Communication; Dynamic; |
| **08** | noun.body | Object; Natural; | **33** | verb.competition | Social; Dynamic; |
| **09** | noun.cognition | Mental; | **34** | verb.consumption | Physical;Usage; Dynamic; |
| **10** | noun.communication | Communication; | **35** | verb.contact | Location; Physical |
| **11** | noun.event | Dynamic; | **36** | verb.creation | Existence; BoundedEvent; |
| **12** | noun.feeling | Experience; | **37** | verb.emotion | Experience; Mental; |
| **13** | noun.food | Comestible; | **38** | verb.motion | Location; Physical; Dynamic; |
| **14** | noun.group | Group; | **39** | verb.perception | Experience; Physical; Dynamic; |
| **15** | noun.location | Place; | **40** | verb.possession | Possession; Dynamic; |
| **16** | noun.motive | 3rdOrderEntity; | **41** | verb.social | Social; Dynamic; |
| **17** | noun.object | Object; | **42** | verb.stative | Static; |
| **18** | noun.person | Human; | **43** | verb.weather | Phenomenal; Physical; Dynamic; |
| **19** | noun.phenomenon | Phenomenal; | | | |
| **20** | noun.plant | Plant; | | | |
| **21** | noun.possession | Possession; | | | |
| **22** | noun.process | Dynamic; | | | |
| **23** | noun.quantity | Quantity; | | | |
| **24** | noun.relation | Relation; | | | |
| **25** | noun.shape | Physical; | | | |
| **26** | noun.state | Static; | | | |
| **27** | noun.substance | Substance; | | | |
| **28** | noun.time | Time; | | | |

The effects of the above heuristics are shown in the table 11 below. We took a random sample of nouns and verbs and measured the quality of the matching by scoring how often the highest match was correct, the 2nd highest match, etc.. This has been done 1) with minimal manual encoding of equivalence relations and applying the tree-matching algorithm, and next after taking each of the above measures in sequence to each result: 2) encoding dubious translations and important synsets by hand and after that running the tree-matching algorithm again, 3) applying the reverse translation option using the English-Dutch dictionary, 4) applying the top-concept matching. In the table, the rows indicate rank of the correct match: the first row the number of times the highest match was correct (match 1), the 2nd correct, the 3rd, 4th, 5th and higher. Obviously, in the ideal case the highest match (rank 1) should be correct. The next row indicates the number of synsets that cannot be translated (presumably a gap in English), which can be translated but there correct translation was not present (non ok) or only a hyperonym translation is given (hyper). Finally, the number of synsets without a translation have been given. The columns then give the improvements, where the first column gives the figures and percentages with minimal manual encoding of equivalence relations (only the Base Concepts), the second column the results after the manual revision of dubious translations, the third column the results of making use of reversed translations and the fourth column the results of matching the top-concepts. The improvements are applied in a cascade. The final column gives the total gain with respect to the first column.

*Table 11: Results of automatic matching heuristics*

| | ***Nouns*** | | | | | | | | | ***Verbs*** | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Rank of correct match** | Tree-match before manual improv. | | Tree-match after manual improv. | | Reversed translation | | Top-Concept Matching | | Gain | Tree-match before manual improv. | | Tree-Match after manual improv. | | Reversed translation | | Top-Concept Matching | | Gain |
| **1** | 60 | 64,5% | 70 | 66,6% | 72 | 68,5% | 74 | 70,4% | 5,9% | 28 | 33,3% | 28 | 32,9% | 32 | 37,6% | 39 | 45,8% | 12,5% |
| **2** | 8 | 8,6% | 10 | 9,5% | 8 | 7,6% | 9 | 8,5% | -0,03% | 7 | 8,3% | 14 | 16,4% | 18 | 21,1% | 12 | 14,1% | 5,7% |
| **3** | 4 | 4,3% | 3 | 2,8% | 3 | 2,8% | 2 | 1,9% | -2,4% | 9 | 10,7% | 10 | 11,7% | 4 | 4,7% | 5 | 5,8% | -4,8% |
| **4** | 0 | 0,0% | 2 | 1,9% | 3 | 2,8% | 1 | 0,9% | 0,9% | 2 | 2,3% | 3 | 3,5% | 4 | 4,7% | 5 | 5,8% | 3,5% |
| **5** | 2 | 2,1% | 2 | 1,9% | 1 | 0,9% | 1 | 0,9% | -1,2% | 1 | 1,1% | 4 | 4,7% | 3 | 3,5% | 4 | 4,7% | 3,5% |
| **>5** | 1 | 1,0% | 1 | 0,9% | 2 | 1,9% | 2 | 1,9% | 0,8% | 0 | 0,0% | 6 | 7,0% | 4 | 4,7% | 2 | 2,3% | 2,3% |
| **gap** | 6 | 6,4% | 9 | 8,5% | 10 | 9,5% | 10 | 9,5% | 3,0% | 8 | 9,5% | 10 | 11,7% | 12 | 14,1% | 10 | 11,7% | 2,2% |
| **all wrong** | 9 | 9,6% | 3 | 2,8% | 4 | 3,8% | 4 | 3,8% | -5,8% | 19 | 22,6% | 6 | 7,0% | 4 | 4,7% | 4 | 4,7% | -17,9% |
| **hyper** | 3 | 3,2% | 5 | 4,7% | 2 | 1,9% | 2 | 1,9% | -1,3% | 4 | 4,7% | 4 | 4,7% | 4 | 4,7% | 4 | 4,7% | -0,06% |
| **Subtotal** | 93 | | 105 | | 105 | | 105 | | | 78 | | 85 | | 85 | | 85 | | 0,0% |
| **notrans** | 102 | 52,3% | 90 | 46,1% | 90 | 46,1% | 90 | 46,1% | -6,1% | 6 | 7,1% | 0 | 0,0% | 0 | 0,0% | 0 | | -7,1% |
| **Total** | 195 | | 195 | | 195 | | 195 | | | 84 | | 85 | | 85 | | 85 | | |
| **Top-3** | 72 | 77,4% | 83 | 79,0% | 83 | 79,0% | 85 | 80,9% | 3,5% | 44 | 52,3% | 52 | 61,1% | 54 | 63,5% | 56 | 65,8% | 13,5% |

If we look at the first match (the highest matching score, 1) we see that for nouns each technique results in about 2% improvement. In total we gained 6% with respect to the first column. In the case of the verbs, we see that the tree-matching has not improved after the manual revision. This is expected because the general effect of tree-matching is poor for verbs. However, the reversed translation technique and, especially, the top-concept matching has resulted in a considerable improvement, 5% and 8% respectively. The total improvement for verbs is therefore even higher than for nouns: 12,55%. Obviously, an increase of correct first matches leads to a decrease of the lower matches.[2]

Since the quality corresponds with the number of matches, we have evaluated the quality per number of matches. In the case of synsets with 1 automatically derived equivalent, 86% of nominal synsets and 78% of the verbal synsets got a correct translation. If there are 2 automatically derived translations, the correct translation was scored as the best match in 68% of the noun synsets and 71% of the verb synsets. In the case of 3-9 translations, the percentages go down to 65% and 49% for nouns and verbs respectively. This figure gets lower the more matches are left, where 10 or more translations are extremely unreliable for verbs (23%). In many of these cases, we are dealing with gaps which cannot properly be translated, even manually. The next table shows the projection of this confidentiality rate to the full set of the automatically extracted equivalences:

[2] Both for verbs and nouns, the number of gaps and the number of translated synsets has increased due to the fact that more word have been translated by the measures explained previously.

*Table 12: Quality of the equivalence relations*

| | *Noun synsets* | | | *Verb synsets* | | |
|---|---|---|---|---|---|---|
| **1** | 9900 | 47.31% | 86% | 569 | 13.82% | 78% |
| **2** | 4246 | 20.29% | 68% | 497 | 12.07% | 71% |
| **3 to 9** | 4673 | 22.33% | 65% | 1907 | 46.32% | 49% |
| **10plus** | 2105 | 10.06% | 54% | 1144 | 27.79% | 23% |
| **Total synsets** | 20924 | | | 4117 | | |

This table shows that 67% of the nouns and 26% of the verbs with automatic translations have 1 or 2 translations with a reasonable reliability (70-86% correct). For verbs still a large proportion has 3 up to 9 translations. For all the cases with 2 up to 9 automatic translations we have selected the top-2 translations. This resulted in 17,838 equivalence links for nouns and 4,808 equivalence links for verbs. Most of these have been checked manually, where we removed 1627 translations for nouns (9% of all 2upto9 links) and 1842 translations for verbs (40% of the 2upto9links). In some cases, we manually added other relations, in other cases, no equivalent could be provided (39 noun synsets, and 107 verb synsets).

If there are 10 or more translations, we decided to neglect the proposed translations. In those cases that the hyperonym in the Dutch wordnet was manually translated, we automatically generated an EQ_HYPERONYM relation or parent-equivalent to the translation of this hyperonym. If there are multiple hyperonyms or multiple equivalents of the hyporonym this may lead to multiple EQ_HYPERONYM relations as well, e.g.:

```
steunpilaar
   has_hyperonym: steun-1
        eq_near_synonym: support {bears the weight of another thing}3149538
        eq_near_synonym: support {holds up or provides a foundation}3150440
   has_hyperonym: pilaar-1
        eq_near_synonym: pillar-1 {a tall cylindrical vertical upright}2326166
```

The Dutch word "steunpilaar" has two hyperonyms in the Dutch wordnet: "steun" (support) and "pilaar" (pillar). No equivalent was generated via the bilingual dictionary and we therefore checked the status of the equivalence relations of the hyperonyms "steun-1" and "pilaar-1". Since the hyperonyms have an eq_(near)_synonym relation which is either manually assigned or marked as OKAY, these are extracted as parent-equivalences to "steunpilaar":

```
steunpilaar
   eq_has_hyperonym: support {bears the weight of another thing}3149538
   eq_has_hyperonym: support {holds up or provides a foundation}3150440
   eq_has_hyperonym: pillar-1 {a tall cylindrical vertical upright}2326166
```

To summarise, the next table gives an overview of manual and automatic equivalence relations for nouns and verbs:

*Table 13: Overview of the equivalence relation status in the Dutch wordnet*

| *Equivalence relation status* | *Source Code* | *Number of equivalence relations for Nouns* | | *Number of equivalence relations for Verbs* | | *Total* | |
|---|---|---|---|---|---|---|---|
| **manual** | 1001 | 3108 | 7.96% | 2203 | 19.94% | 5311 | 10.61% |
| **manual and okay** | 1002 | 697 | 1.79% | 265 | 2.40% | 962 | 1.92% |
| **heuristics and okay** | 1003 | 4791 | 12.28% | 3685 | 33.36% | 8476 | 16.93% |
| **heuristics** | 1004 | 26432 | 67.73% | 3531 | 31.97% | 29963 | 59.84% |
| **automatic parent-equivalent** | 1005 | 3995 | 10.24% | 1362 | 12.33% | 5357 | 10.70% |
| **Total** | | 39023 | 100.00% | 11046 | 100.00% | 50069 | 100.00% |
| **Manually verified** | | 8596 | 22.03% | 6153 | 55.70% | 14749 | 29.46% |

First of all, we see that 22% of the noun equivalences and 56% of the verb equivalences is somehow manually verified (the sum of rows 1, 2 and 3). Between 10-12% of the equivalences is a parent-equivalent that is automatically generated because the heuristics generated 10 or more candidates. The remaining cases are generated by the heuristics, where 9,900 nominal synsets have only one translation (86% reliability) and 569 verbal synsets (78% reliability). For the remaining cases, we have manually removed 10% of the nominal links and 40% of the verbal links that were wrong. The column with source code refers to the number code that identifies the source of the equivalence relations in the text file of the Dutch wordnet (in EuroWordNet format). In section 4.3, we will further explain how the different status of the equivalence relations is encoded in the database files.

## *3.3. The creation of the Dutch core wordnet*

As explained at the beginning of this section, most of the effort has been devoted to the Dutch core wordnet, built around the Base Concepts. The manual work for this has mostly been carried out in the ALS database. Special editors and query options have been developed to easily create links between synsets or sets of synsets. Queries can be applied incrementally so that precise selections of synsets can be made that share particular features (e.g. morphological, orthographic, semantic, or corpus frequency). The following information has been used for building the core wordnets:

1. queries over definitions patterns
2. morphology
3. bilingual dictionaries
4. comparison with WordNet1.5 structures and synsets

We will first describe how the selection of the Base Concepts has been established and next what procedures have been followed to encode and restructure the relations.

### *3.3.1 Selection of Dutch Base Concepts and equivalences for the Common Base Concepts*

The Base Concepts are the starting point for building the core wordnets. The selection should include the most generic concepts of a language that still comprehensively represent the diversity of the complete vocabulary. As explained in section 2, the Vlis database does not form a *closed* hierarchy, in which all relations are unified in a single tree or a small set of tops, but it has 1429 noun tops and 298 verb tops. We therefore did not just take the Vlis tops as a starting point for selecting the base concept but we used the following combination of criteria:

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

We first selected nominal concepts with more than 15 relations (965 synsets) and all verbal concepts with more than 12 relations (237 synsets). These selections represent about 15% of all the relations in the database. We added to these selections all hyperonyms of these concepts that not have been included. This resulted in a first set of 1165 noun and 285 verb senses:

*Table 14: Base Concept selection based on the number of relations*

| | *15% of all relations* | *hyperonyms* | *union* |
|---|---|---|---|
| **NOUNS with >15 relations** | 965 | 470 | 1165 |
| **VERBS with >12 relations** | 237 | 120 | 285 |

Inspection of this set showed that some of these concepts occur at rather specific levels. The second criterion was therefore to look at the position in the hierarchy. We limited the initial selection to those concepts that occur within the first 2 levels of the Vlis Database, which turned out to be the most generic cut off point. This limited the first set to 706 nouns and 268 verbs, respectively. The next column shows how many of above concepts occur in the first 3 levels.

*Table 15: Base Concept selection distributed over the 3 highest levels of the hierarchy*

| *Levels* | *Nouns* | | *Verbs* | |
|---|---|---|---|---|
| | intersection (number of senses) | percentage of the first selection (1165 senses) | intersection (number of senses) | percentage of the first selection (1165) |
| **level 0** | 148 | 12,7% | 94 | 32,9% |
| **down to level 1** | 432 | 37,1% | 209 | 73,3% |
| **down to level 2** | **706** | **60,6%** | **268** | **94,0%** |
| **down to level 3** | 921 | 79,0% | 276 | 96,8% |

The above procedure excluded some important concepts or areas of the vocabulary because none of the concepts in the hierarchy had substantial relations. It is possible that whole branches of the hierarchy are interrelated via many levels with a relatively low number of relations. On the other hand, we did not wanted to add all the tops from the Vlis database. We therefore added those tops or ends that have more than 10 children at any depth of the hierarchy. The extension is shown in the next table:

*Table 16: Non-selected hierarchy tops with more than 10 children*

| | *Tops* | *>10 children at any depth* | *Not yet included* | *Extended Selection* |
|---|---|---|---|---|
| **NOUNS** | 1429 | 114 | 35 | 741 |
| **VERBS** | 298 | 50 | 14 | 223 |

Finally, we looked at word frequency information in the Celex lexicon. Inspection of all words with a frequency of 2000 or more showed that frequency is not a sufficient criterion for genericity of concepts. The set includes both very general words, such as "meubel" (*furniture*) and "fruit" (*fruit*) but also very specific words at the Basic Level, "bed" (*bed*) and "appel" (*apple*). We therefore used the frequency criterion only as a filter on both the number of relations and the Vlis tops. Nouns with more than 15 and verbs with more than 12 relations, or Vlis tops

that were excluded by the above procedure, but had a word frequency according to Celex of 2000 or more occurrences have been added:

*Table 17: Non-selected concepts which are tops or have many relations but with high corpus frequency*

| | *Initital Selection* | *Celex Frequency > 2000 Vlis End >12 verbal relations >15 nominal relations* | *Final Selection of Dutch Base Concepts* |
|---|---|---|---|
| **NOUNS** | 741 | 164 | 905 |
| **VERBS** | 223 | 61 | 284 |

As explained in the general wordnet documentation, this set has been translated manually to the closest WordNet1.5 synsets and the translations have been used for the selection of the common Base Concepts. The Common Base Concepts (1300 concepts in total) include concepts that are not represented in the above Dutch set and there are Dutch Base Concepts (DBCs) that did not make it into the set of Common Base Concepts (CBCs). This is shown in the next table:

*Table 18: Dutch Base Concepts (DBC) compared to the Common Base Concepts (CBC) in EuroWordNet*

| | *Proposed DBC* | *Selected DBC* | *Rejected DBC* | *Missing CBC* |
|---|---|---|---|---|
| **NOUNS** | 1027 | 429 | 598 | 265 |
| **VERBS** | 323 | 126 | 197 | 51 |
| **TOTAL** | 1350 | 555 | 795 | 316 |

To make sure that the Common Base Concepts are well represented in the Dutch wordnet, we thus had to extend our selection with Dutch equivalences for the 'missing' concepts in this table. In some cases (97 in total), there was no equivalent for the Common Base Concepts in the Dutch wordnet. We then created a so-called complex equivalence link to the closest concept(s), as is illustrated in the following examples:

| **Common Base Concept** | **Equivalence relation** | **Closest Dutch Concept** |
|---|---|---|
| cause to feel unwell#1, Verb | EQ_IS_CAUSED_BY | {onwel#1}, Adjective (sick) |
| vessel#2, Noun | EQ_INVOLVED | {bevatten#1}, Verb, (to contain) |
| | EQ_HYPERONYM | {doos#1}, Noun, (box) |
| | EQ_HYPERONYM | {zak#1}, Noun, (bag) |
| | EQ_HYPERONYM | {blik#1}, Noun, (tin) |
| | EQ_HYPERONYM | {kist#1}, Noun, (box) |

Both the synset {cause to feel unwell} and {vessel, container} do not exist as lexicalized forms in Dutch. In the case of {cause to feel unwell}, the closest Dutch concept is the adjective "onwel" (unwell, sick), which can be related as the result of this Base Concept (BC). In the case of {vessel, container}, the closest equivalents are the verb equivalent of *to contain* and some more specific *containers* in Dutch that are lexicalized. The BC can thus be represented by multiple complex links to these concepts. Note that all these Dutch synsets also have direct equivalences in English (*sick*, *to contain*, *box*, *bag*, *tin*). The above relations are thus additional to the direct equivalents, only to have a precise mapping of the BCs to Dutch.

In total about 1200 Dutch concepts have been related to the CBCs. Together with the 795 Dutch synsets that have not been selected as Common BCs but that do play an important role in the Dutch wordnet (according to the above criteria), they constituted a core set of about 2,000 Dutch word meanings that have been used to build the Dutch core wordnet.

*3.3.2. Manual encoding of the language internal and equivalence relations*

The 2,000 Dutch Base Concepts, (more than 50% of which represent Common Base Concepts, also encoded in the other wordnets) have been processed according to the following general procedure:

1. select closely related terms and word meanings;
2. establish the major semantic classes and differentiation;
3. provide a hyperonym classification to the top of the hierarchy;
4. chart out hierarchical differences among the selected words;
5. establish the correct equivalence relations for the most important concepts;
6. establish any other language-internal relation in so far necessary;

In many cases, there is a set of closely related words that roughly cover the semantic space of a Base Concept. In some cases, these words are listed as synonyms, sometimes they have hyponymy relations, but in other cases they are unrelated (e.g. because one of them is a top in the hierarchy or because they have been classified according to different perspectives). Below these words we may find various clusters of hyponyms. The first step to be taken is to make a comprehensive list of these words. Several techniques are available for finding these word clusters:

- co-occurrence data from corpora;
- expanding from closely related words and synsets in WordNet1.5;
- word meanings with similar definitions, one-word-definitions, circular definitions;
- overlapping translations in bilingual dictionaries;

The first technique is well-know in information theory. Words that tend to have the same co-occurrence pattern also tend to be similar in meaning (Sparck-Jones and Willett 1997). This can be applied to independent corpora or to the definitions themselves.

The second technique is rather obvious. By directly translating the synset members in WordNet1.5 it is possible to derive synsets in another language. This can be extended to hyponyms. In some cases, the translations of these synsets or their hyponyms give rise to other classes and categories that have not been thought of or have not been included in first Dutch set.

The third technique looks at definitions that are very similar, and, in particular, definitions consisting of a single word or circularly defining words in terms of each other. This is illustrated by the following Dutch examples:

| | |
|---|---|
| *apparaat* | min of meer samengesteld **werktuig** |
| (apparatus) | (more or less assembled tool) |
| *instrument* | min of meer samengesteld of fijn **gereedschap** of **toestel**… |
| (instrument) | (more or less assembled or delicate tool or apparatus) |
| *toestel* | **apparaat** |
| (apparatus) | (apparatus) |
| *werktuig* | stuk **gereedschap** |
| (tool) | (piece of tools) |
| *gereedschap* | **werktuig** |
| (tools, instruments) | (tool) |

Here we see 5 different meanings that are circularly defined, suggesting a synonymy relation.

Another possibility is to look for words that have the same translations and/or occur as translations for the same words in bilingual dictionaries. The procedure is more or less the same as the translation heuristic that used the bilingual dictionaries. Starting with a set of closely related Dutch words extracted on the basis of other techniques, such as the previous instrument examples *apparaat* (apparatus), *toestel* (apparatus), and *werktuig* (tool), and *gereedschap*(tools), we extract all the English translations for all their meanings from the bilingual Dutch-English dictionary. Next all these English translations are looked up in the reverse English-Dutch dictionary to see what Dutch words are given as translations for all the different meanings. The result is a very large list of translation-sets, covering very different meanings. However, we keep only those sets of Dutch translations that include at least two of the original words with which the search was started. These sets form a so-called translation-cycle via two bilingual resources (note that any language-pair can be used for this). The co-occurrence of pairs of source words is thus used as a filter to select the correct meaning of the word. The automatically-generated result for the above words is the following list:

**Potential Equivalents generated from bilingual dictionaries:**

| | |
|---|---|
| *gebruiksvoorwerp* 1 | (implement, appliance, utensil) |
| *comfort* 1 | (comfort) |
| *mechanisme* 2 | (mechanism) |
| *inrichting* 5 | (construction, installation) |
| *tuig* 1 | (gear, equipment) |
| *uitmonstering* 3 | (equipment, outfit, kit) |
| *uitrusting* 1 | (equipment) |
| *outillage* 1 | (equipment) |
| *apparatuur* 1 | (apparatus, machinery) |
| *materieel* 1 | (material, equipment) |
| *machinerie* 1 | (machinery) |
| *systeem* 10 | (system) |
| *mechaniek* 1 | (mechanism) |

All these words express some kind of instrumentality. Among them are a few synonyms but also words that can be related in other ways.

After establishing a reasonable set of related words and word clusters for a Base Concept or Base Concept representatives a careful study is made of the hyponyms distributed over the different classes, possible levels of the hyponyms and the hyperonym relations to the top of the hierarchy. The following main variations tend to occur in hierarchical classifications (Vossen 1995):

- Similar words are classified at different levels of abstraction;
- Different but more-or-less equivalent words have been used to classify the same meanings;
- Other perspectives have been chosen to classify similar meanings;

In the next hierarchy (Figure-3) containing Dutch words for diseases we see a typical combination of the phenomena, where multiple perspectives and levels have been missed:

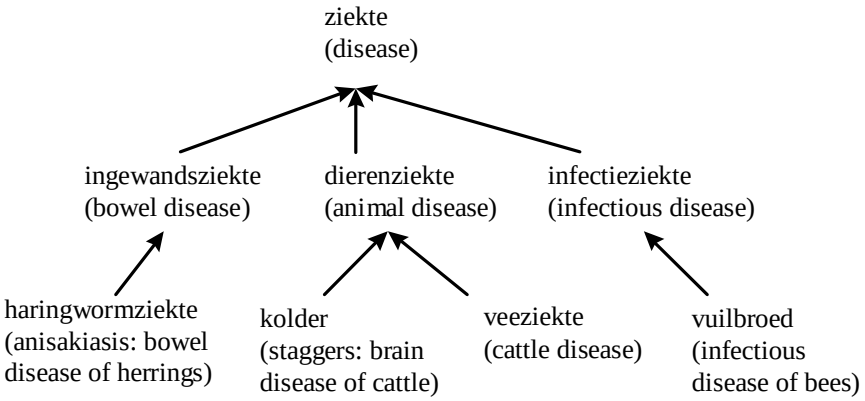


*Figure 3: Hierarchical Relations in the Van Dale database*

We see here that *haringwormziekte* (anisakiasis) is only linked to *ingewandziekte* (bowel disease) and that *vuilbroed* (infectious disease of bees) is only linked to *infectieziekte* (infectious disease), while both are diseases of animals: *herrings* and *bees* respectively. In both cases, the classification as *dierenziekte* (animal disease) has been omitted. Within the same part of the hierarchy we see the opposite situation for *kolder* (staggers) which is directly linked to *dierenziekte* (animal disease) while it is also a disease of *cattle* and should be linked to *veeziekte* (cattle disease). One of the reasons for the incompleteness of the classifications is the lack of multiple hyperonyms in the Vlis hierarchy.

The hierarchy of diseases contains some typical examples of restructuring that are required because sub-levels of hyperonyms have been skipped and multiple classifications have been missed. Such variation in levels and multiple classifications can be detected by applying the Principle-of-Economy to the hyponyms (Dik 1978). This principle states that it is not allowed to relate a word W1 to a word W3 when there is a word W2 linked to W3 to which W1 can be linked in the same way. In practice this means that all hyponyms of *ziekte* (disease) have to be

cross-checked to see whether they represent hyperonyms of each other.[3] This then also reveals multiple category membership. In EuroWordNet, multiple hyperonyms have been encoded more systematically. When applied to the above cases, we get the following structure:

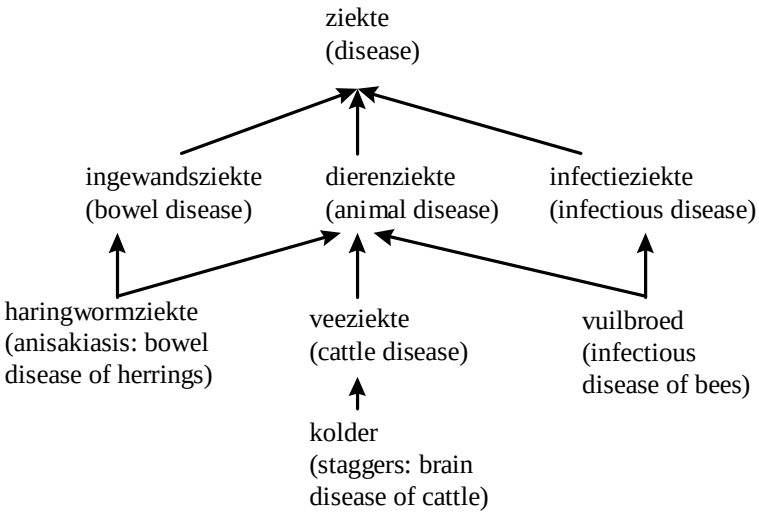


*Figure 4: Restructured Hierarchical Relations in the Dutch wordnet*

A further special effort has been made to build up the hyperonym chains from Base Concepts such as "disease" to link them to a unified top-level that is compatible to the EWN top-ontology, possibly using multiple hyperonyms.

Other changes involved:

- the splitting or merging of synsets;
- the reallocation of hyponyms after the major classes have been established;
- the specification of translations for major classes;

Where necessary, Base Concepts or closely related concepts have been further enriched with non-hyponymy relations.

The Dutch core wordnet consists of about 10,000 synsets related to the 2,000 most generic Base Concepts. This set has been compared with the other wordnets in terms of coverage and compatibility of relations. The results of the comparison have been used to extend and improve the wordnet.

[3] Some practical strategies for finding similar meanings which are classified differently, is by making use of the morphology of the entries (e.g. compounds ending with *disease*), or by looking for other, alternative definition patterns (e.g. containing phrases such as *infectious*).

### *3.4. Extension of the core wordnet to the complete Dutch wordnet*

In the second building phase for the Dutch wordnet, we focussed on:

1. extending the core wordnet to the full size;
2. improving the overlap across the wordnets;
3. improving the quality of the equivalence relations;
4. verify the corpus frequency of the selection;
5. include regular lexicalization patterns;
6. adding new spelling variants;

The work on the equivalence relations has been described in section 3.2. above. The other extensions/improvements are discussed here.

The core wordnet for Dutch covered 9,588 synsets: 5,917 nominal synsets, 3,282 verbal synsets and 389 adjectival/adverbial synsets. This core wordnet had to be extended to approximately 35K synsets (25K nominal synsets and 10K verbal synsets) or 50K word meanings (more than 1 synonym on average per synset).

For the verbs this implied extending the set to the complete lexicon of 9,125 synsets made available by Van Dale. These synsets contain 14,278 senses and 8,868 entries. The verbal wordnet has 100% overlap with the Parole lexicon for Dutch for entries with a frequency above 100.[4] The verbal wordnet for Dutch has two verbal tops "zijn" (to be) for static verbs and "gebeuren" (to happen) for dynamic verbs. All verbal synsets are connected to one of these tops via at least one hyperonym link. In addition, they may have other links (see tables in section 4).

The coverage of the nominal wordnet has been increased by several measures:

1. Extending classes with significantly low coverage of ontological clustering, as followed from the top-concept clustering in WordNet1.5. This mainly involved 1stOrderEntities;
2. Investigation of significant gaps in the Dutch wordnet based on a comparison of coverage of the Inter-Lingual-Index (ILI). These are translations of hierarchy nodes covered by the other wordnets but not in the Dutch wordnet;
3. Extending the vocabulary with missing Parole entries (all senses) with a frequency above 100;
4. Adding new spelling variants to the synsets;
5. Inclusion of all synsets with 10 or more relations;
6. Inclusion of all direct hyponyms (1 level) of concepts with 50 or more relations;
7. Inclusion of all synsets with 1 or 2 automatically derived translations;
8. Inclusion of al hyperonyms needed to classify the above concepts;

Coverage of the First Subset has been measured by clustering of synsets per Top Concept. This clustering is achieved by collecting the Top Concepts for all translations to WordNet1.5 synsets

[4] The Dutch Parole lexicon is developed by the Instituut voor Nederlandse Lexicografie (INL). The lexicon contains morpho-syntactic information for the most frequent words taken from corpora (Kruyt 1998). The INL has compared the Dutch wordnet entries with their lexicon for different frequency clusters to measure the overlap.

or hyperonyms of these translations in WordNet (see D029D030 (Vossen et al. 1999) on the comparison of the wordnets). Since the aimed size of the final wordnet is 1/3rd of the size of WordNet1.5, the ideal balancing is also 1/3rd of the coverage of WordNet1.5. for each cluster.

The comparison of the 2ndOrderEntity clusters (events, states, relations, processes) did not show any unbalanced clusters across the sites. No balancing of the coverage was needed. In fact, relatively many 2ndOrderEntities are encoded, when compared to WordNet1.5. In the case of the 1stOrderEntities the following clusters had significantly lower coverage in the Dutch wordnet compared to the coverage of WordNet1.5: Animal; Creature; Function; Garment; Gas; Group; Human; Living; Occupation; Plant; Software. Except for Group, we have maximally balanced the clusters by adding more specific concepts. In the case of Animal and Plant, it is however impossible to get the same distribution because many exotic classes in WordNet1.5 are not available in our resource. However, as it appeared that many common *animal* and *plant* names were missing in our database, many of these have been added to the database (801 synsets in total).

An overall comparison of the intersection of the wordnets (where concepts are represented by the ILI-records with which they have an equivalence mapping) generated feedback on major hierarchy nodes that have been covered by other languages (English, Spanish and Italian) but not in Dutch. We investigated the most important nodes in the other wordnets which are lacking in the Dutch wordnet and improved our subset where possible. The following explanations have been found for the gaps:

1. there is no equivalent:
    1.1. because it does not exist in Dutch: it is a genuine gap.
    1.2. it does exist but was not taken up in our resource:
        1.2.1. for non explicit reasons
        1.2.2. because it would be a multi-word (which we did not include in our wordnet)
2. there is an equivalent:
    2.1. but we disagreed with the classification, therefore no changes were made
    2.2. we agree with the classification, but somehow it was not assigned (yet), so we adapted the classification.

One of the important conclusions from this comparison is that there are some major differences between the Dutch hierarchy and the WordNet1.5. hierarchy. Some important classes high up the hierarchy are treated differently, such as: "relation", "meaning", "communication", "thought", "area". This has important consequences for the comparison of the wordnets.

To verify that frequent words are represented in the Dutch wordnet, we have compared the Dutch wordnet with the most frequent words from the Parole project. There are 6492 nominal entries in the database that have a frequency above 100 in the Parole corpus. These correspond with 7945 noun senses in the Vlis database. All these senses have been included in the Dutch wordnet. In section 4, we give a table that compares the vocabulary of the Dutch wordnet with the vocabulary of Celex, distributed over the Celex frequency.

The entries in the Vlis database are all in the Dutch spelling from before 1997. In 1997, a new spelling has been agreed. We have therefore expanded the synsets in the Dutch wordnet with new

spelling variants. For this purpose we used a word form list that has been built for a Dutch-Russian dictionary that is developed at the University of Amsterdam. This word form list includes new spelling forms. We first made a selection of all forms that could not be found in the Dutch wordnet. Next, we modified the unmatched entries by undoing the spelling changes. Entries that could be matched after the spelling change was reversed, have been added to the matching entries in their changed (non-matching) form as new spelling variants. For example, one change involves the addition of a binding "n" in Dutch compounds. The old spelling contained words such as "kattevoer" (cat food), which are now spelled as "kattenvoer", adding a binding "n". The form "kattenvoer" cannot be matched with the Dutch wordnet, but by removing the "n" we can retrieve the entry "kattevoer". In that case "kattenvoer" is added as a new variant to all senses of "kattevoer". In total 896 new nominal variants have been added using this procedure. Also a few verbal (10) and adjectival (6) variants have been added.

Next we made sure that the selection of additional concepts was distributed over the major semantic field and included the concepts for which reliable information could be derived. To make sure that major fields are well-represented we have selected:

- all concepts with more than 10 relations;
- all direct hyponyms of words with 50 or more relations;

The first measure ensured that more important concepts are present, while the second measure leads to more comprehensive lists of concepts just below major concepts. Just taking all direct hyponyms of the Base Concepts would lead to too many concepts.

Furthermore, to make sure that all concepts with reliable information have been included we selected:

- all concepts that had at least 1 manually encoded or modified relation (either language-internal or equivalent relation);
- all concepts with only 1 or 2 automatically derived equivalence relations (these have the highest reliability);

Finally, we excluded concepts that had no translation, no relations except a hyperonym, a single meaning and a low frequency. In addition, we generated all the hyperonym chains for the selected concepts to verify that also the hyperonyms are part of the selection.

The next table shows how the final nominal subset has been compiled.

*Table 19: Selection of the nominal synsets for the Dutch wordnet*

| *Selection criterion* | *Number of Synsets* |
|---|---|
| all synsets which have a manually processed relation (language internal or equivalence) | 15,364 |
| all synsets with only 1 automatically extracted equivalence relation | 9747 |
| all synsets with only 2 automatically extracted equivalence relations | 4173 |
| all synsets with 10 or more relations | 1637 |
| all direct hyponyms (1 level) of synsets with 50 or more relations | 10,766 |
| all senses of Parole entries with a frequency above 100 (6492 entries) | 7945 |
| Synsets with new spelling variants | 846 |
| Manually added animal/plant synsets | 801 |
| all hyperonyms (all levels) of the above union | 5272 |
| Total | **34,455** |

Obviously, the sets selected by these criteria overlap considerably. The total set of 34,455 synsets is therefore not the sum of these figures. All these nominal synsets are related via hyponymy-links to a single top: "iets" (anything).

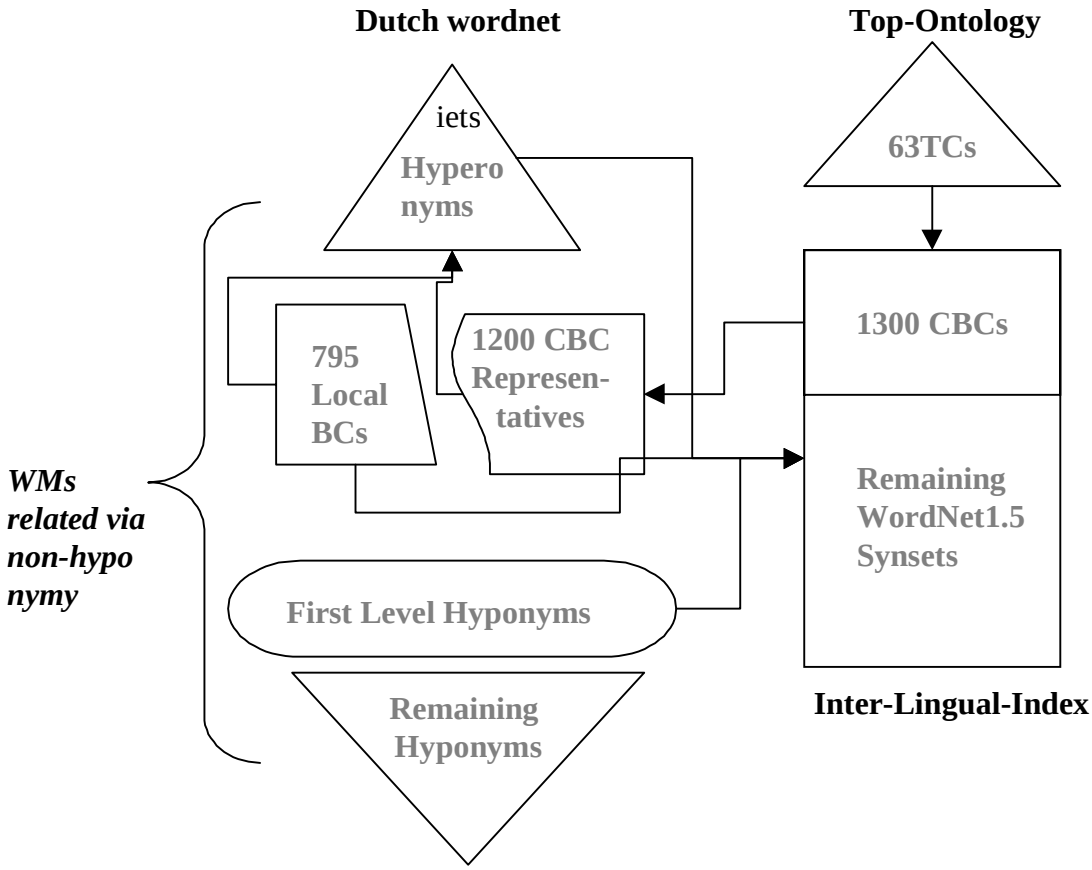


*Figure 5: Overview of the vocabulary of the Dutch wordnet*

Except for the animal-plant synsets and the new spelling variants, all entries and senses originate from the Vlis database. This is shown in the next table:

*Table 20: Origin of the variants in the Dutch wordnet*

| *Var Source* | *Source Code* | *Nouns* | | *Verbs* | | *Others* | |
|---|---|---|---|---|---|---|---|
| **Origin is Vlis** | 2001 | 52731 | 96.88% | 14141 | 99.93% | 1616 | 99.63% |
| **New Animal/Plant** | 2002 | 801 | 1.47% | 0 | 0.00% | 0 | 0.00% |
| **New Spelling Variant** | 2003 | 896 | 1.65% | 10 | 0.07% | 6 | 0.37% |
| **Total** | | 54428 | | 14151 | | 1622 | |

Again the source code is used to mark the origin of synset variants in the text file of the Dutch wordnet.

Manual processing of the full extension was not possible, given the limitations in time and budget. We therefore used some overall strategies to detect deficiencies:

- overall checks on the structural coherence
- unifying the hyperonym relations in a single top-node
- encoding of non-hyponymy relations for large semantic fields or clusters
- checking of the compatibility of relations with top-concept classifications

To get an overview of the conceptual coherence, we exported the complete hierarchy as sorted flat chains with ALS:

iets 1 middel 2 vervoermiddel 1 voertuig 1 motorrijtuig 1 auto 1 automaat 3
iets 1 middel 2 vervoermiddel 1 voertuig 1 motorrijtuig 1 auto 1 bedrijfsauto 1
iets 1 middel 2 vervoermiddel 1 voertuig 1 motorrijtuig 1 auto 1 begrafenisauto 1
iets 1 middel 2 vervoermiddel 1 voertuig 1 motorrijtuig 1 auto 1 bezemwagen 1
iets 1 middel 2 vervoermiddel 1 voertuig 1 motorrijtuig 1 auto 1 brandweerauto 1

Incoherent collections of leaves, tops, cycles and mistakes at the higher levels (at the left side) can easily be detected in this way. After we made sure that large and crucial nodes in the Dutch hierarchy are acceptable, we replaced this chain by the WordNet1.5 equivalents if available. If there is not translation we maintained the Dutch word. The result for the above example looks as follows:

iets 1 means 1 conveyance 3 vehicle 1 automotive vehicle 1 auto 1 automatic 1
iets 1 means 1 conveyance 3 vehicle 1 automotive vehicle 1 auto 1 bedrijfsauto 1
iets 1 means 1 conveyance 3 vehicle 1 automotive vehicle 1 auto 1 hearse 1
iets 1 means 1 conveyance 3 vehicle 1 automotive vehicle 1 auto 1 bezemwagen 1
iets 1 means 1 conveyance 3 vehicle 1 automotive vehicle 1 auto 1 fire engine 1

Because the Dutch hierarchy was already checked, incoherences pointed to wrong or bad translations, e.g.:

'chemical compound_1' --> 'building_1'
mind_7 -> group_1

All odd cases have been inspected and the translations have been corrected when necessary. This option also makes it possible to verify if drastic mistakes have been made at major points in the hierarchy.

A separate effort has been made to differentiate large clusters of hyponyms using non-hyponymy relations. Typical clusters are:

- persons, differentiated by their agent/patient_roles or co_roles;
- manners of motion, differentiated by the involved instruments, the location or direction;
- creation events, differentiated by the involved_result;
- processing of materials, differentiated by the patient_role and the instrument_role;
- instruments, differentiated by their instrument_role or co_instrument_patient;
- parts and groups, differentiated by the meronymy relations;

Finally, we have checked the constraints on the semantic relations. As explained in the general EuroWordNet document, there are all kinds of semantic limitations on the target and source concepts of a relation (in addition to the part-of-speech limitations). A distinction can be made between type-persistent relations and type-shifting relations. Type-persistent relations only hold between entities of the same type: either 1stOrderEntities (concrete things), 2ndOrderEntities (events, processes, states and relations) or 3rdOrderEntities (mental objects). Examples of type-persistent relations are: hyponymy, meronymy, cause, subevent, antonymy, xpos_synonymy. Type-shifting relations are: roles/involvements, be_in_state. By selecting major hyponymy nodes in the Dutch hierarchy (e.g. "voorwerp" (object)), it is possible to search the ALS database for senses within the scope of that hierarchy in which particular types of relations occur. This functionality has been used to check whether relations are compatible with the type constraints, or relations have been reversed by accident.

The above measures have lead to structural and quantitative improvements to the Dutch hierarchy. The next table gives an overview of the origin of the language-internal relations for the final Dutch wordnet:

*Table 21: Origin of the relations in the Dutch wordnet*

| *Relation Source* | *Source Code* | *Nouns* | | *Verbs* | | *Others* | | *Total* | |
|---|---|---|---|---|---|---|---|---|---|
| **Copied from Vlis and Okay** | 1 | 17140 | 20,20% | 7500 | 28,77% | 4 | 0,49% | 24644 | 22,05% |
| **Copied from Vlis** | 2 | 43614 | 51,39% | 8466 | 32,47% | 2 | 0,25% | 52082 | 46,61% |
| **Manual and Okay** | 101 | 18336 | 21,61% | 8872 | 34,03% | 570 | 70,46% | 27778 | 24,86% |
| **Manual** | 102 | 5779 | 6,81% | 1233 | 4,73% | 233 | 28,80% | 7245 | 6,48% |
| **Total** | | 84869 | | 26071 | | 809 | | 111749 | |
| **Manual Total** | | 41255 | 48,61% | 17605 | 67,53% | 807 | 99,75% | 59667 | 53,39% |

As before, the source code indicates the origin of the language-internal relations in the database. Overall, 51% of the noun relations and 32% of the verb relations directly originate from Vlis. A further 20% and 28% of the nouns and verbs, respectively, is copied from Vlis and manually checked in EuroWordNet. Finally, 28% of the nouns and 38% of the verbs is added manually in EuroWordNet. This table does not indicate how many relations have been removed.

The column with *Others* represents the adjectives and adverbs. Due to lack of an agreed model or specification for relations between adjectives, hardly any relations have been taken over from Vlis. All relations are encoded manually, as the result of coding the nouns and verbs that are closely related to these adjectives. No relations have been encoded between adjectives.

# 4. The structure and content of the Dutch wordnet

We will first give some quantitative overview tables and in the next subsection we will explain some typical structures of the Dutch wordnet.

## *4.1. Quantitative overview*

The next table specifies the number of synsets, word senses, words and the total number relations and equivalence relations for the Dutch wordnet. For comparison, the same figures are given for WordNet1.5 (except for the equivalence relations).

*Table 22: Numbers of synsets, senses, entries and relations in the Dutch wordnet and WordNet1.5*

| | *Dutch Wordnet* | | | | *WordNet1.5* | | | |
|---|---|---|---|---|---|---|---|---|
| | *Noun* | *Verb* | *Oth* | *Total* | *Noun* | *Verb* | *Oth* | *Total* |
| **Synsets** | 34455 | 9040 | 532 | 44015 | 60521 | 11363 | 22631 | 94515 |
| **No. of senses** | 54428 | 14151 | 1634 | 70201 | 107428 | 25768 | 54406 | 187602 |
| **Sens./syns.** | 1.58 | 1.57 | 3,07 | 1.59 | 1,78 | 2,27 | 2,40 | 1,98 |
| **Entries** | 45972 | 8826 | 1495 | 56283 | 88175 | 14734 | 23708 | 126617 |
| **Sens./entry** | 1.18 | 1.60 | 1,09 | 1.25 | 1,22 | 1,75 | 2,29 | 1,48 |
| **LIRels.** | 84869 | 26071 | 833 | 110711 | 159975 | 24332 | 27821 | 212128 |
| **LIRels/syns** | 2,46 | 2,88 | 1,57 | 2,52 | 2,64 | 2,14 | 1,23 | 2,24 |
| **EQRels-ILI** | 42055 | 23077 | n.a. | 65132 | | | | |
| **EQRels/syn** | 1.22 | 2.55 | n.a. | 1.48 | | | | |
| **Synsets without ILI** | 6070 | 1133 | n.a. | | | | | |

The first table shows that the Dutch wordnet exceeds the minimally aimed size of 35K synsets and 50K wordnets. For nouns, 56,8% of the size of WordNet1.5 has been reached and in the case of verbs even 79,5%. The ratios of senses per synset (degree of synonymy) and senses per entry (degree of polysemy) are both a bit lower in the Dutch wordnet. Note however that the Dutch wordnet also makes use of near_synonymy as a relation between synsets, which lowers the degree of synonymy. If we look at the language-internal relations (LIRs), we see that comparable numbers of relations have been expressed as in WordNet1.5 (a bit less for nouns and a bit more for verbs). In general, we can say that a rich and condensed lexicons with a minimal levels of polysemy is a good property of a database. The next table shows how the relations are distributed over the different relation types.

Obviously, no comparison of the equivalence relations can be made with WordNet1.5. Nevertheless, figures close to 1 equivalent on average can be considered to be good. This is the case for nouns but not for verbs. As can be in Table 23 below, these are mostly due to generalization equivalents that make up 51% of all the verbal equivalence relations. These are automatically generated by the database for clustered sense-groups. (see general EuroWordNet document for an explanation). Only a small proportion of the Dutch synsets did not receive an equivalent link.

*Table 23: Distribution of the language-internal relations in the Dutch wordnet*

| *Language Internal Relations* | *Nouns* | | *Verbs* | | *Other* | | *Total* | |
|---|---|---|---|---|---|---|---|---|
| **HAS_HYPERONYM** | 35861 | 42,3% | 10160 | 39,0% | | | 46021 | 41,6% |
| **HAS_HYPONYM** | 35861 | 42,3% | 10160 | 39,0% | | | 46021 | 41,6% |
| **HAS_XPOS_HYPERONYM** | 29 | 0,0% | 51 | 0,2% | 29 | 3,5% | 109 | 0,1% |
| **HAS_XPOS_HYPONYM** | 56 | 0,1% | 28 | 0,1% | 1 | 0,1% | 85 | 0,1% |
| **NEAR_SYNONYM** | 182 | 0,2% | 104 | 0,4% | | | 286 | 0,3% |
| **XPOS_NEAR_SYNONYM** | 896 | 1,1% | 763 | 2,9% | 180 | 22,2% | 1839 | 1,7% |
| **NEAR_ANTONYM** | 288 | 0,3% | 388 | 1,5% | | 0,0% | 676 | 0,6% |
| **XPOS_NEAR_ANTONYM** | 3 | 0,0% | 3 | 0,0% | | 0,0% | 6 | 0,0% |
| **HAS_HOLONYM** | 260 | 0,3% | | | | | 260 | 0,2% |
| **HAS_HOLO_LOCATION** | 258 | 0,3% | | | | | 258 | 0,2% |
| **HAS_HOLO_MADEOF** | 143 | 0,2% | | | | | 143 | 0,1% |
| **HAS_HOLO_MEMBER** | 220 | 0,3% | | | | | 220 | 0,2% |
| **HAS_HOLO_PART** | 1135 | 1,3% | | | | | 1135 | 1,0% |
| **HAS_HOLO_PORTION** | 67 | 0,1% | | | | | 67 | 0,1% |
| **HAS_MERONYM** | 260 | 0,3% | | | | | 260 | 0,2% |
| **HAS_MERO_LOCATION** | 258 | 0,3% | | | | | 258 | 0,2% |
| **HAS_MERO_MADEOF** | 143 | 0,2% | | | | | 143 | 0,1% |
| **HAS_MERO_MEMBER** | 220 | 0,3% | | | | | 220 | 0,2% |
| **HAS_MERO_PART** | 1135 | 1,3% | | | | | 1135 | 1,0% |
| **HAS_MERO_PORTION** | 67 | 0,1% | | | | | 67 | 0,1% |
| **INVOLVED** | 62 | 0,1% | 47 | 0,2% | | 0,0% | 109 | 0,1% |
| **INVOLVED_AGENT** | 213 | 0,3% | 472 | 1,8% | | 0,0% | 685 | 0,6% |
| **INVOLVED_PATIENT** | 384 | 0,5% | 591 | 2,3% | | 0,0% | 975 | 0,9% |
| **INVOLVED_INSTRUMENT** | 287 | 0,3% | 1263 | 4,8% | | 0,0% | 1550 | 1,4% |
| **INVOLVED_LOCATION** | 89 | 0,1% | 77 | 0,3% | | 0,0% | 166 | 0,1% |
| **INVOLVED_RESULT** | 55 | 0,1% | 234 | 0,9% | | 0,0% | 289 | 0,3% |
| **INVOLVED_DIRECTION** | 0 | 0,0% | 9 | 0,0% | | 0,0% | 9 | 0,0% |
| **INVOLVED_SOURCE_DIRECTION** | 4 | 0,0% | 10 | 0,0% | | 0,0% | 14 | 0,0% |
| **INVOLVED_TARGET_DIRECTION** | 2 | 0,0% | 27 | 0,1% | | 0,0% | 29 | 0,0% |
| **ROLE** | 109 | 0,1% | | | | | 109 | 0,1% |
| **ROLE_AGENT** | 685 | 0,8% | | | | | 685 | 0,6% |
| **ROLE_PATIENT** | 975 | 1,1% | | | | | 975 | 0,9% |
| **ROLE_INSTRUMENT** | 1550 | 1,8% | | | | | 1550 | 1,4% |
| **ROLE_LOCATION** | 166 | 0,2% | | | | | 166 | 0,1% |
| **ROLE_RESULT** | 289 | 0,3% | | | | | 289 | 0,3% |
| **ROLE_DIRECTION** | 9 | 0,0% | | | | | 9 | 0,0% |
| **ROLE_SOURCE_DIRECTION** | 13 | 0,0% | | | 1 | 0,1% | 14 | 0,0% |
| **ROLE_TARGET_DIRECTION** | 25 | 0,0% | | | 4 | 0,5% | 29 | 0,0% |
| **CO_ROLE** | 42 | 0,0% | | | | | | |
| **CO_AGENT_PATIENT** | 41 | 0,0% | | | | | | |
| **CO_AGENT_INSTRUMENT** | 49 | 0,1% | | | | | | |
| **CO_AGENT_RESULT** | 42 | 0,0% | | | | | | |
| **CO_PATIENT_AGENT** | 41 | 0,0% | | | | | | |
| **CO_PATIENT_INSTRUMENT** | 282 | 0,3% | | | | | | |
| **CO_PATIENT_RESULT** | 1 | 0,0% | | | | | | |
| **CO_INSTRUMENT_AGENT** | 49 | 0,1% | | | | | | |
| **CO_INSTRUMENT_PATIENT** | 282 | 0,3% | | | | | | |
| **CO_INSTRUMENT_RESULT** | 83 | 0,1% | | | | | | |
| **CO_RESULT_AGENT** | 42 | 0,0% | | | | | | |
| **CO_RESULT_PATIENT** | 1 | 0,0% | | | | | | |
| **CO_RESULT_INSTRUMENT** | 83 | 0,1% | | | | | | |
| **CAUSES** | 306 | 0,4% | 885 | 3,4% | | | 1191 | 1,1% |
| **IS_CAUSED_BY** | 405 | 0,5% | 383 | 1,5% | 403 | 49,8% | 1191 | 1,1% |
| **HAS_SUBEVENT** | 225 | 0,3% | 185 | 0,7% | | | 410 | 0,4% |
| **IS_SUBEVENT_OF** | 190 | 0,2% | 220 | 0,8% | | | 410 | 0,4% |
| **MANNER_OF** | 15 | 0,0% | | | 2 | 0,2% | 17 | 0,0% |
| **IN_MANNER** | 14 | 0,0% | 3 | 0,0% | | | 17 | 0,0% |
| **BE_IN_STATE** | 313 | 0,4% | | | | | 313 | 0,3% |
| **STATE_OF** | 97 | 0,1% | 8 | 0,0% | 208 | 25,7% | 313 | 0,3% |
| **FUZZYNYM** | 2 | 0,0% | | | | | 2 | 0,0% |
| **XPOS_FUZZYNYM** | 5 | 0,0% | | | 5 | 0,6% | 10 | 0,0% |
| **Total** | 84869 | | 26071 | | 833 | | 110735 | |
| **Synsets** | 34455 | | 9040 | | 532 | | 44027 | |
| **Average per synset** | 2,46 | | 2,88 | | 1,57 | | 2,52 | |

Table 23 shows that hyponymy (hyperonyms and hyponyms) is the most important language-internal relation: 78% for nouns and 83,2% for verbs. Another important relation is XPOS_NEAR_SYNONYM, which connects nouns, verbs and adjectives that can be used to refer to the same entities. Relations between different entity types are the role/involved relations. If we take the subtotals of the specific subtypes together, we still get a considerable number of relations: for nouns 1096 involved relations (1,3%) and 3821 role relations (4,5%), for verbs 2730 involved relations (10,5%). Other relations for nouns are holonym and meronym (2083 relations or 2,5% each), and for verbs causal relations (1259 or 4,9%). In the case of adjectives/adverbs, we see that only non-hyponymy relations are encoded, all of which are across part-of-speech.

The next table gives the distribution of the equivalence relations over the different relation types:

*Table 24: Distribution of the equivalence relations in the Dutch wordnet*

| *Equivalence Relations* | *Nouns* | | *Verbs* | | *Total* | |
|---|---|---|---|---|---|---|
| **EQ_SYNONYM** | 2006 | 4.77% | 312 | 1.35% | 2318 | 3.56% |
| **EQ_NEAR_SYNONYM** | 31625 | 75.20% | 8552 | 37.06% | 40177 | 61.69% |
| **EQ_HAS_HYPERONYM** | 4796 | 11.40% | 2217 | 9.61% | 7013 | 10.77% |
| **EQ_HAS_HYPONYM** | 191 | 0.45% | 32 | 0.14% | 223 | 0.34% |
| **EQ_INVOLVED** | 17 | 0.04% | 34 | 0.15% | 51 | 0.08% |
| **EQ_ROLE** | 59 | 0.14% | | | 59 | 0.09% |
| **EQ_IS_CAUSED_BY** | 15 | 0.04% | 29 | 0.13% | 44 | 0.07% |
| **EQ_CAUSES** | 7 | 0.02% | 32 | 0.14% | 39 | 0.06% |
| **EQ_HAS_HOLONYM** | 89 | 0.21% | | | 89 | 0.14% |
| **EQ_HAS_MERONYM** | 37 | 0.09% | | | 37 | 0.06% |
| **EQ_HAS_SUBEVENT** | 2 | 0.00% | 6 | 0.03% | 8 | 0.01% |
| **EQ_IS_SUBEVENT_OF** | | | 7 | 0.03% | 7 | 0.01% |
| **EQ_BE_IN_STATE** | 22 | 0.05% | 3 | 0.01% | 25 | 0.04% |
| **EQ_IS_STATE_OF** | 4 | 0.01% | | | 4 | 0.01% |
| **EQ_CO_ROLE** | 22 | 0.05% | | | 22 | 0.03% |
| **EQ_GENERALIZATION** | 2762 | 6.57% | 11786 | 51.07% | 14548 | 22.34% |
| **EQ_METONYM** | 401 | 0.95% | 17 | 0.07% | 418 | 0.64% |
| **EQ_DIATHESIS** | | | 50 | 0.22% | 50 | 0.08% |
| **Total** | 42055 | | 23077 | | 65132 | |

The average number of ILIs per synset is 1.56 for nouns and 3.34 for verbs, not considering the synsets without ILIs. There are 6,070 nominal synsets (16%) and 1,133 verbal synsets (12%) that have no ILI reference. We see that EQ_NEAR_SYNONYM is the most frequent relation (75% for nouns and 37% for verbs). These are mostly due to the automatic procedures to derive equivalence relations. Also the EQ_GENERALIZATION, EQ_METONYM and EQ_DIATHESIS relations are derived automatically. They are added by the EuroWordNet database to each Dutch synset that is related to a member of a so-called ILI-cluster or group. ILI-clusters have been added to the database as Composite ILI-records to group closely related meanings (see the general EuroWordNet documentation). As said above, EQ_GENERALIZATION makes up 51% of all verbal equivalence relations. All other relations are added manually, either to represent the Base Concepts or to correct wrongly-generated automatic relations. Except for EQ_SYNONYMS, most of these manual links indicate that the Dutch concept does not exist in WordNet1.5.

As explained above, all words from the Vlis database with a frequency above 100 in the Dutch Parole corpus are included in the Dutch wordnet. The next table shows the distribution of the Dutch wordnet vocabulary over the Celex frequencies. We can see here that nouns and verbs with a frequency above 1000 have a coverage of 91-93%. This slowly decreases to about 75% for frequencies above 100. Obviously, words with zero frequency have been covered less. The total coverage is 89% for nouns and 71.5% for verbs. In general we can conclude that the most frequent words are well-represented in the Dutch wordnet.

*Table 25: Overlap and distribution of the Dutch wordnet entries over Celex corpus frequencies*

| | | *Noun Entries* | | | *Verb Entries* | |
|---|---|---|---|---|---|---|
| **Frequency** | **Celex** | **Dutch Wordnet** | **Coverage** | **Celex** | **Dutch Wordnet** | **Coverage** |
| **1001+** | 1217 | 1130 | 92.85% | 677 | 620 | 91.58% |
| **501-1000** | 939 | 831 | 88.50% | 455 | 398 | 87.47% |
| **251-500** | 1408 | 1176 | 83.52% | 637 | 516 | 81.00% |
| **101-250** | 3157 | 2346 | 74.31% | 1176 | 889 | 75.60% |
| **51-100** | 3604 | 2302 | 63.87% | 957 | 684 | 71.47% |
| **31-50** | 3380 | 1918 | 56.75% | 695 | 500 | 71.94% |
| **21-30** | 3016 | 1607 | 53.28% | 495 | 316 | 63.84% |
| **11-20** | 5258 | 2631 | 50.04% | 706 | 473 | 67.00% |
| **6-10** | 4804 | 2195 | 45.69% | 567 | 360 | 63.49% |
| **3-5** | 4713 | 2079 | 44.11% | 377 | 247 | 65.52% |
| **2** | 2338 | 946 | 40.46% | 346 | 203 | 58.67% |
| **1** | 127 | 68 | 53.54% | 9 | 6 | 66.67% |
| **0** | 30001 | 10641 | 35.47% | 1725 | 1096 | 63.54% |
| **Total** | 33565 | 29870 | 88.99% | 8822 | 6308 | 71.50% |

## *4.2. Major structures and clusters*

The complete Dutch wordnet has two top nodes: "iets" (anything) and “niets” (nothing), which are related by antonymy. Below “niets” we find nothing, below “iets” we find all other concepts in the Dutch wordnet. All nominal synsets (except “niets) are related to “iets” via hyponymy-links. All verbal concepts are linked to two verbal tops "gebeuren" (to happen) and "zijn" (to be), which are also linked to “iets” via cross-part-of-speech hyponymy. The pronoun "iets" can substitute any phrase in Dutch, including verbs and verb phrases and thus represents a natural top of the wordnet.

The indented list below gives the first level of hyponyms below "iets" in the Dutch wordnet. There are two types of children:

- LEAVES that do not have hyponyms at a further level;
- NODES with hyponyms at deeper levels;

NODES are followed by two figures, the first of which indicates the direct children below it, and the second node indicates the total number of children at any level. We have marked the most important NODES here. These NODES are also followed by an English translation and proceeded by a label that indicates the kind of entities that it classifies, where 1st = 1stOrderEntities, 2nd = 2ndOrderEntities and 3rd = 3rdOrderEntities. In some cases, the NODE cuts across these entity types, which is indicated by multiple classifications.

```
_ iets 1
                HAS_HYPONYM Allerheiligste 1 LEAF
                HAS_HYPONYM aanrader 1 NODE 1 / 1
                HAS_HYPONYM achterste 3 NODE 1 / 1
1st             HAS_HYPONYM afscheiding 2 NODE 19 / 232 (separation)
                HAS_HYPONYM attractie 1 NODE 5 / 12
                HAS_HYPONYM bedreiging 2 NODE 2 / 2
                HAS_HYPONYM belemmering 2 NODE 4 / 17
                HAS_HYPONYM blikvanger 1 LEAF
1st             HAS_HYPONYM bron 2 NODE 12 / 126 (source)
1st;2nd;3rd     HAS_HYPONYM deel 2 NODE 375 / 5691 (part)
                HAS_HYPONYM desideratum 1 LEAF
                HAS_HYPONYM dinges 1 LEAF
                HAS_HYPONYM eigendom 2 LEAF
2nd             HAS_HYPONYM eigenschap 1 NODE 103 / 4231 (property)
                HAS_HYPONYM entiteit 1 LEAF
                HAS_HYPONYM equivalent 1 LEAF
2nd             HAS_HYPONYM gebeurtenis 1 NODE 52 / 13198 (event, NOUN)
3rd             HAS_HYPONYM gedachte 2 NODE 30 / 1368 (thought)
1st;2nd;3rd     HAS_HYPONYM geheel 1 NODE 29 / 629 (whole)
                HAS_HYPONYM geheim 1 NODE 7 / 7
                HAS_HYPONYM gemak 2 NODE 1 / 1
1st;2nd;3rd     HAS_HYPONYM groep 4 NODE 102 / 5923 (group)
                HAS_HYPONYM gruwel 1 LEAF
1st             HAS_HYPONYM hoeveelheid 1 NODE 128 / 768 (quantity)
                HAS_HYPONYM hybride 2 LEAF
                HAS_HYPONYM inhoud 2 NODE 5 / 62
                HAS_HYPONYM investering 1 LEAF
                HAS_HYPONYM juweel 2 NODE 2 / 2
                HAS_HYPONYM kneus 2 LEAF
                HAS_HYPONYM lokkertje 1 NODE 2 / 2
1st;2nd;3rd     HAS_HYPONYM middel 2 NODE 50 / 5266 (means)
                HAS_HYPONYM namaak 1 NODE 9 / 12
                HAS_HYPONYM nieuwigheid 1 NODE 2 / 2
1st             HAS_HYPONYM object 1 NODE 4 / 17986 (object)
                HAS_HYPONYM onderpand 2 LEAF
                HAS_HYPONYM onding 1 LEAF
                HAS_HYPONYM oorsprong 1 NODE 6 / 15
                HAS_HYPONYM oorzaak 1 NODE 11 / 28
                HAS_HYPONYM oppepper 1 LEAF
                HAS_HYPONYM overblijfsel 1 NODE 6 / 16
                HAS_HYPONYM parel 2 LEAF
1st             HAS_HYPONYM plaats 1 NODE 92 / 4041 (place)
                HAS_HYPONYM prestige-object 1 LEAF
1st             HAS_HYPONYM produkt 1 NODE 48 / 2469 (product)
                HAS_HYPONYM puikje 1 LEAF
                HAS_HYPONYM rommel 3 NODE 1 / 1
1st             HAS_HYPONYM samenstelling 3 NODE 1 / 441 (composition)
                HAS_HYPONYM schande 2 LEAF
1st             HAS_HYPONYM soort 2 NODE 29 / 227 (kind)
                HAS_HYPONYM stimulans 2 NODE 2 / 3
1st             HAS_HYPONYM substantie 1 NODE 89 / 4993 (substance)
                HAS_HYPONYM succes 2 NODE 1 / 2
2nd             HAS_HYPONYM tijd 1 NODE 10 / 824 (time)
2nd             HAS_HYPONYM toestand 1 NODE 123 / 4329 (situation, NOUN)
                HAS_HYPONYM trots 3 NODE 1 / 1
1st;2nd         HAS_HYPONYM uiting 2 NODE 41 / 5975 (utterance)
                HAS_HYPONYM uitslag 1 LEAF
                HAS_HYPONYM uitvinding 1 LEAF
                HAS_HYPONYM versiering 1 NODE 31 / 56
                HAS_HYPONYM vuil 1 NODE 1 / 1
                HAS_HYPONYM woekering 1 NODE 1 / 3
                HAS_HYPONYM zaak 1 LEAF
2nd             XPOS_HAS_HYPONYM gebeuren 2 NODE (to happen, VERB)
2nd             XPOS_HAS_HYPONYM zijn 7 NODE (to be, VERB)
```

Even though, the NODES are most important because they establish the major semantic structures, the LEAVES are interesting from a linguistic point of view. A LEAF at this level of the hierarchy is almost void. It can be applied to any type of entity (it is not restricted) and expresses some kind of role, function or state (see Vossen 1995), e.g.:

| | |
|---|---|
| oorzaak (cause) | a person, thing, idea or event that is the cause of an event |
| overblijfsel (remainder) | a person, thing, idea or event that is the remains of an event |
| eigendom (property) | a person or thing that is owned |

Here the disjunction in the definition indicates that the words can be used for an open range of entities. Other cases have a more limited usage, possibly because of the kind of property or function that they express. Words at high levels of the hierarchy and open denotations are called functionals.

We see that also some of the major NODES can be called functionals. Words such as "deel" (part), "middel" (means), "groep" (group), are orthogonal with the classification of 1stOrder, 2ndOrder and 3rdOrderEntities, e.g.:

| | |
|---|---|
| 1st | "een deel van zijn hoofd" (part of his head) |
| 2nd | "een deel van de vergadering" (part of the meeting) |
| 3rd: | "een deel van deze theorie" (part of this theory) |

| | |
|---|---|
| 1st | "deze pil is een middel tegen kanker "<br>(this pill is a 'means' against cancer) |
| 2nd | "vergaderen is een belangrijk middel tot communiceren"<br>(to have meetings is an important means of communication) |
| 3rd: | "dit idee is een middel voor het oplossen van het probleem<br>(this idea is a means to solve the problem) |

The lower we get, the more specific concepts become and thus less orthogonal. Below "deel" we find many concrete parts (e.g. body parts, parts of machines, buildings) and member-group relations, and below "means" we find all instruments.

Other concepts below "iets" represent more homogeneous categories of more specific concepts:

- 1stOrderEntities are found below "object" (all objects, both animate and inanimate, natural and artifact), and below "substantie" (substances, natural and artificial);
- 2ndOrderEntities are found below the nouns "gebeurtenis" (event), "eigenschap" (property), "toestand" (situation) and the verbs "zijn" (to be) and "gebeuren" (to happen);
- 3rdOrderEntities, such as "idee" (idea), "concept" (concept), "kennis" (knowledge), "plan" (plan), can be found below "gedachte" (thought);

Obviously, these categories cross-classify with many other. Objects (below "object") will typically cross-classify with places (below "plaats"), parts ("deel"), instruments ("middel"), products ("produkt"), whereas substances will typically cross-classify with products ("produkt"), compositions ("samenstelling"), quantity ("hoeveelheid"). Cross-classification leads to multiple hyperonym schemes.

Below we see the indented lists for the major verb clusters “gebeuren” (to happen) and “zijn” (to be), structured in the same way as the first level of “iets” (anything) above:

```
_ gebeuren 2
   HAS_HYPONYM beginnen 3 NODE 8 / 12
   HAS_HYPONYM doorgaan 4 NODE 3 / 4
   HAS_HYPONYM gebeuren 3 LEAF
   HAS_HYPONYM handelen 2 NODE 53 / 4429 (to act)
   HAS_HYPONYM herhalen 3 LEAF
   HAS_HYPONYM meemaken 1 NODE 9 / 256 (to experience)
   HAS_HYPONYM missen 4 NODE 2 / 2
   HAS_HYPONYM overkomen 5 LEAF
   HAS_HYPONYM plaatsvinden 1 NODE 4 / 5
   HAS_HYPONYM samenvallen 2 NODE 1 / 1
   HAS_HYPONYM stoppen 8 NODE 10 / 21
   HAS_HYPONYM terugkomen 2 LEAF
   HAS_HYPONYM uitoefenen 2 NODE 4 / 421 (to excel force)
   HAS_HYPONYM veranderen 1 NODE 154 / 2358 (to change)
   HAS_HYPONYM verkopen 1 LEAF
   HAS_HYPONYM verliezen 1 NODE 8 / 9
   HAS_HYPONYM verlopen 4 NODE 2 / 2
   HAS_HYPONYM veroorzaken 1 NODE 48 / 6242 (to cause)
   HAS_HYPONYM vinden 1 NODE 5 / 11
   HAS_HYPONYM voorafgaan 1 LEAF
   HAS_HYPONYM voorkomen 4 NODE 6 / 10
   HAS_HYPONYM wachten 2 LEAF


_ zijn 7
   HAS_HYPONYM achterliggen 2 NODE 1 / 1
   HAS_HYPONYM afhangen 2 LEAF
   HAS_HYPONYM bedreigen 2 LEAF
   HAS_HYPONYM betekenen 2 NODE 2 / 3
   HAS_HYPONYM bevinden 2 NODE 27 / 255
   HAS_HYPONYM blijven 5 NODE 7 / 17
   HAS_HYPONYM denken 4 NODE 9 / 53
   HAS_HYPONYM dichtzitten 1 LEAF
   HAS_HYPONYM domineren 1 LEAF
   HAS_HYPONYM doorlaten 1 NODE 4 / 6
   HAS_HYPONYM eruitzien 1 NODE 13 / 39
   HAS_HYPONYM gaan 15 LEAF
   HAS_HYPONYM horen 1 NODE 6 / 8
   HAS_HYPONYM houden 3 LEAF
   HAS_HYPONYM innemen 2 NODE 4 / 16
   HAS_HYPONYM koken 1 NODE 1 / 1
   HAS_HYPONYM kunnen 2 NODE 6 / 8
   HAS_HYPONYM kunnen 4 NODE 2 / 2
   HAS_HYPONYM leiden 2 NODE 7 / 8
   HAS_HYPONYM leiden 5 NODE 1 / 1
   HAS_HYPONYM lijken 1 NODE 4 / 4
   HAS_HYPONYM lijken 2 NODE 8 / 10
   HAS_HYPONYM moeten 4 NODE 1 / 4
   HAS_HYPONYM onderdoen 1 LEAF
   HAS_HYPONYM ontgaan 2 LEAF
   HAS_HYPONYM overeenkomen 1 NODE 7 / 12
   HAS_HYPONYM plegen 2 LEAF
   HAS_HYPONYM schemeren 1 NODE 1 / 1
   HAS_HYPONYM verschillen 1 NODE 6 / 8
   HAS_HYPONYM weergeven 2 NODE 2 / 18
   HAS_HYPONYM weten 2 NODE 5 / 11
   HAS_HYPONYM zijn 2 NODE 5 / 21
   HAS_HYPONYM zullen 2 NODE 1 / 1
```

The dynamic cluster “gebeuren” (to happen) contains several fundamental classes:

| | |
|---|---|
| causal change: | veroorzaken (to cause); |
| non-causal change: | veranderen (to change as a process) |
| phenomena: | voorkomen (to occur); plaatsvinden (to take place); overkomen (to happen) |
| actions: | handelen (to act) |
| experience: | meemaken (to undergo, experience) |
| exert force: | uitoefenen (to exert force) |
| aspectuals: | beginnen (to begin); doorgaan (to continue); stoppen (to stop); herhalen (to repeat); samenvallen (to co-occur); |

The causal changes below “veroorzaken” (to cause) and the non-causal processes below “veranderen” (to change as a process) represent the largest clusters. Below these, we find, among others, all movement verbs, existential verbs (*make*, *kill*, *create*, *produce, etc.*), and verbs of physical change (*soften*, *harden*, *damage, etc.*). The causal meanings are often transitive, the non-causal mostly intransitive. A special group are phenomena that can be found below “voorkomen” (to occur), "plaatsvinden" (to take place); "overkomen" (to happen). These take place or occur (mostly with a natural cause) and without a necessarily clear result. Below “handelen” (to act), we find many human acts, including subclasses such as *change of possession*, *speech acts*, *thinking*. The experience cluster below “meemaken”, contains *feelings* and *emotions*, at least involving an experiencer and possibly a stimulus. Below “uitoefenen” (to exert force), we find classes like “duwen” (push), “trekken” (pull), “zuigen” (suck), “raken” (hit). These are further differentiated either in terms of the *result*, *causality* or *instrumentality*, possibly by combining them with other categories. A special category are the aspectuals. There are not many genuine aspectuals but aspectual verbs are often combined with other more specific verbs of change (see below).

An important feature of the verb hierarchy is the use of multiple hyperonyms to capture regular compositional verb structures. In Dutch, it is possible to derive composite verbs such as the following:

| | |
|---|---|
| doorademen: | door+ademen (lit. through+breath, continue to breath) |
| doorbetalen: | door+betalen (lit. through+pay, continue to pay) |
| doorlopen: | door+lopen (lit. through+walk, continue to walk) |
| doorfietsen: | door+fietsen (lit. through+walk, continue to walk) |
| doorrijden: | door+rijden (lit. through+walk, continue to walk) |

The first set of examples contains compounds combining the particle “door” (though) with a main verb, which has an aspectual semantic effect: to continue the event expressed by the main verb.

The aspectual effect is also expressed by one of the above generic aspectual verbs:

| | |
|---|---|
| doorgaan: | door+gaan (lit. through+go, continue); |

This means that we can systematically apply multiple hyponymy to these compound verbs, capturing both the aspectual implication and the event expressed by the main verb:

| | |
|---|---|
| doorademen: | door+ademen (lit. through+breath, continue to breath) |
| HAS_HYPERONYM | ademen (breath) |
| HAS_HYPERONYM | doorgaan (continue) |
| doorbetalen: | door+betalen (lit. through+pay, continue to pay) |
| HAS_HYPERONYM | betalen (pay) |
| HAS_HYPERONYM | doorgaan (continue) |

A similar phenomena can also be observed for compound verbs in which a main verb is preceded by an adective/adverb that denotes the resulting state:

| | |
|---|---|
| openknijpen: | open+knijpen (lit. open+squeeze, to open by squeezing) |
| opendraaien: | open+draaien (lit. open+turn, to open by turning) |
| openschuiven: | open+schuiven (lit. open+slide, top open by sliding) |
| dichtknijpen: | dicht+knijpen (lit. closed+squeeze, to close by squeezing) |
| dichtdraaien: | dicht+draaien (lit. closed +turn, to close by turning) |
| dichtschuiven: | dicht+schuiven (lit. closed +slide, to close by sliding) |

In all these cases, there is also a main verb that expresses just the resulting change of state without the main verb:

| | |
|---|---|
| openmaken: | open+maken (lit. open+make, to cause to be open); |
| dichtmaken: | dicht+maken (lit. dicht+make, to cause to be open); |

We can thus again use multiple hyponymy to capture both the result implication and the event expressed by the main verb:

| | |
|---|---|
| openknijpen: | open+knijpen (lit. open+squeeze, to open by squeezing) |
| HAS_HYPERONYM | knijpen (squeeze) |
| HAS_HYPERONYM | openmaken (to open) |
| opendraaien: | open+draaien (lit. open+turn, to open by turning) |
| HAS_HYPERONYM | draaien (to turn) |
| HAS_HYPERONYM | openmaken (to open) |
| dichtknijpen: | dicht+knijpen (lit. closed+squeeze, to close by squeezing) |
| HAS_HYPERONYM | knijpen (squeeze) |
| HAS_HYPERONYM | dichtmaken (to close) |
| dichtdraaien: | dicht+draaien (lit. closed +turn, to close by turning) |
| HAS_HYPERONYM | draaien (to turn) |
| HAS_HYPERONYM | dichtmaken (to close) |

Using multiple hyponymy links to rather basic verbs at several levels, complex meanings can be composed in a relatively systematic way. Note that the above meanings are productive but that in many cases there is also an unproductive meaning in addition to the above word senses in the wordnet.

Another remarkable property of the noun and verb hierarchies is that there appears to be a double system for nouns and verbs denoting 2ndOrderEntities that is quite parallel. At the highest level, we find a parallel division into dynamic events and static situations, both as nominal forms and verbal forms:

dynamic nodes
  NOUN gebeurtenis (event)
  VERB gebeuren (to happen))
static nodes:
  NOUN toestand (situation)
  NOUN eigenschap (property)
  VERB zijn (to be)

The next Figure-6 shows a similar parallelism at deeper levels in the dynamic cluster, where at major, parallel points a XPOS_SYNONYM relation is encoded between the nominal and verbal chain:

| **Nominal Hierarchy** | | | | | **Verbal Hierarchy** | | | |
|---|---|---|---|---|---|---|---|---|
| gebeurtenis (event) | | | | <XPOS_SYN> | | | | gebeuren (to happen) |
| | handeling (an act) | | | <XPOS_SYN> | | | handelen (to act) | |
| | | daad (deed) | | <XPOS_SYN> | | doen (to do) | | |
| | verandering (a change) | | | <XPOS_SYN> | | | veranderen (to change, intrans) | |
| | | beweging (movement) | | <XPOS_SYN> | | bewegen (to move, intrans) | | |
| | | | verplaatsing (change of location) | <XPOS_SYN> | verplaatsen (to move position, intrans) | | | |
| | | | | <XPOS_SYN> | | | veroorzaken (to cause) | |
| | verandering (a change) | | | <XPOS_SYN> | | veranderen (to change, trans) | | |
| | | beweging (movement) | | <XPOS_SYN> | bewegen (to move, trans) | | | |

*Figure 6: Parallel hierarchies for nominalized events and verbs*

There are many of these parallelisms, and they can be exploited to check consistency across the hierarchies or to systematically paraphrase text as noun phrases or sentences. There may be some disturbances due to the fact that the nominalized version abstracts from the argument structure and consequently is less explicit about the causality of the event. We thus see that both the transitive and the intransitive versions of verbs of change (“veranderen”) map to a single

nominalized form: "verandering" (a change). We also see that the causality implication introduces an extra level "veroorzaken" (to cause) in the causative chain.

If we look at the hyponyms of the static verbal top “zijn” (to be) we first of all see that they form a relatively small group compared to the previous clusters. Static situations tend to be lexicalized as nouns and adjectives/adverbs in Dutch. Static nouns (below “eigenschap” (property) and “toestand” (situation)) form a large cluster. They include category names for physical adjectival properties:

kleur (colour); maat (size); volume (volume); lengte (length); smaak (taste); vorm (shape);

and for social and mental states or attitudes, such as:

armoede (poverty); rijkdom (richness); werkeloosheid (unemployment);
gevoel (feeling); twijfel (doubt); geloof (belief); haat (hatred); liefde (love);

In a sense, the nominal static hierarchy covers many different dimensions of the adjectival hierarchy. A complete connection between the two hierarchies could not be made in the Dutch wordnet because there was not sufficient time and resources for properly encoding the relations between adjectives/adverbs. As an illustration, we have currently linked “kleur” (colour), “smaak” (taste) and “vorm” (shape) with the most frequent adjectival values via an has_xpos_hyponym relation, e.g.:

“kleur” (colour)
has_xpos_hyponym: zwart (black), wit (white); rood (red); geel (yellow); blauw (blue); groen (green); paars (purple); oranje (orange)

There is also some parallelism between the static nominals and the hyponyms below “zijn”, but this is less systematic and predictable. For example, “eruitzien” (have a physical appearance) groups many physical properties and “denken-4” (consider in the attitudinal sense) contains mental attitudes. For the rest, there is not much hierarchical structure and only incidental overlap between nouns and verbs. We will find many property denoting verbs in flat layers below “bevinden” (to be positioned in some situation), among which verbs that express spatial positions: “zitten” (to sit), “staan” (to stand), “liggen” (to lay), “hangen” (to hang), and other conditions such as “slapen” (to sleep), “vastzitten” (to be stuck). Another limited cluster is formed by existential “zijn” (to exist).

One interesting group is formed by the modal verbs, most of which are directly linked below “zijn” (to be) and have a few hyponyms:

“kunnen” = to be able to
“moeten” = to be obliged to
“zullen” = to do in the future

Since they are matrix verbs that express some modal aspect with respect to a complement verb phrase, they can be seen as static property-denoting verbs. Although they have a rather technical meaning, there are still some interesting lexical semantic relations that can be encoded. There are

some dynamic verbs such as “laten” (to let), “toestaan” (to allow), “verbieden” (to forbid) that can be seen as acts or speech acts and have a causal relation with the modal properties denoted by the static verbs. Furthermore, “kunnen” (to be able to) is related to “vermogen” (ability) that groups many physical and mental abilities (awareness, intelligence, sight, smell, hearing), which are again related to perceptual values and experiences.

Finally, the flat hierarchical structure of static verbs, adjectives and nouns makes it useful to make a more tight structure via xpos_synonymy relations and to classify them by means of more abstract features like physical, mental, social, as is done using the EuroWordNet top-ontology. Hyponyms in a large cluster such as “eigenschap” (property) denote many different types of properties, ranging from mental, perceptual, physical, modal to relational or social. This information is provided in the glosses of these synsets (see the next section).

The next levels of below “iets” (anything), “zijn” (to be) and “gebeuren” (to happen) contain many other clusters, which are not further shown here. For example, all *animals*, *plants*, *humans* are grouped below "wezen" (being), which is a direct hyponym of "object" (object). In the Appendix we have expanded the 2nd level, but limited the lists to all hyponyms with 10 or more children at any level. This reveals a bit of the further hyponymic structure.

Non-hyponymy relations have been encoded in 3 different cases:

- for major nodes and Base Concepts in the hierarchy
- cause and role/involved relations to differentiate large classes or shallow hierarchies
- to clarify the meaning if there was not a good equivalent in English
- to create some comprehensive thematic grids

The first strategy has been explained extensively above. The second strategy has for example been applied to *instruments* and their purpose or function, to *humans* and their role or properties and to *changes* and their result, e.g.:

| | |
|---|---|
| brandblusser (fire extinguisher) | role_instrument blussen (to extinguish fire) |
| condesator (condenser) | role_instrument opslaan (to store) |
| decoder (decoder) | role_instrument decoderen (to decode) |
| frankeermachine (franking machine) | role_instrument frankeren (to frank) |
| broodrooster (toaster) | role_instrument roosteren (to toast) |
| | co_instrument_patient brood (bread)$_{N}$ |
| shredder (shredder) | role_instrument versnipperen (to shred) |
| | co_instrument_patient paper (paper)$_{N}$ |
| wasdroger (dryer) | role_instrument drogen (to dry) |
| | co_instrument_patient was (cloth to be washed)$_{N}$ |
| | |
| afwezige (absent person)$_{N}$ | be_in_state afwezig (absent)$_{A}$ |
| nieuweling (a new person)$_{N}$ | be_in_state nieuw (new)$_{A}$ |
| verdachte (suspect person)$_{N}$ | be_in_state verdacht (suspect)$_{A}$ |
| bevrijder (liberator)$_{N}$ | role_agent bevrijden (to set free) |
| bewonderaar (admirer)$_{N}$ | role_agent bewonderen (to admire) |
| gokker (gambler)$_{N}$ | role_agent gokken (to gamble) |
| roker (smoker)$_{N}$ | role_agent roken (to smoke) |

| | |
|---|---|
| natmaken (to wet, make wet) | causes nat (wet) $_{A}$ |
| harden (to harden) | causes hard (hard) $_{A}$ |
| gladmaken (to smoothen) | causes glad (even, smooth) $_{A}$ |
| composteren (to compost) | involved_result compost (compost) $_{N}$ |
| creosoteren (to creosoot) | involved_result creosoot (creosoot) $_{N}$ |

In the case of "wasdroger" (dryer), we see that the related concept "was" (clothes that will, are or have been washed) does not exist in English. In such cases, the non-hyponymy relations are very helpful to clarify the meaning.

Because of limitations in time and budget, the fourth strategy has been applied occasionally to particular areas of the vocabulary. Nevertheless, we think it is a very fruitful way of encoding lexical semantic relations. Two areas that have been covered in this way are *sports* and *diseases*. By specifying the exact relations between all concepts related to the theme, a comprehensive and natural coverage of a domain can be achieved. This is illustrated in Figure-7 below:

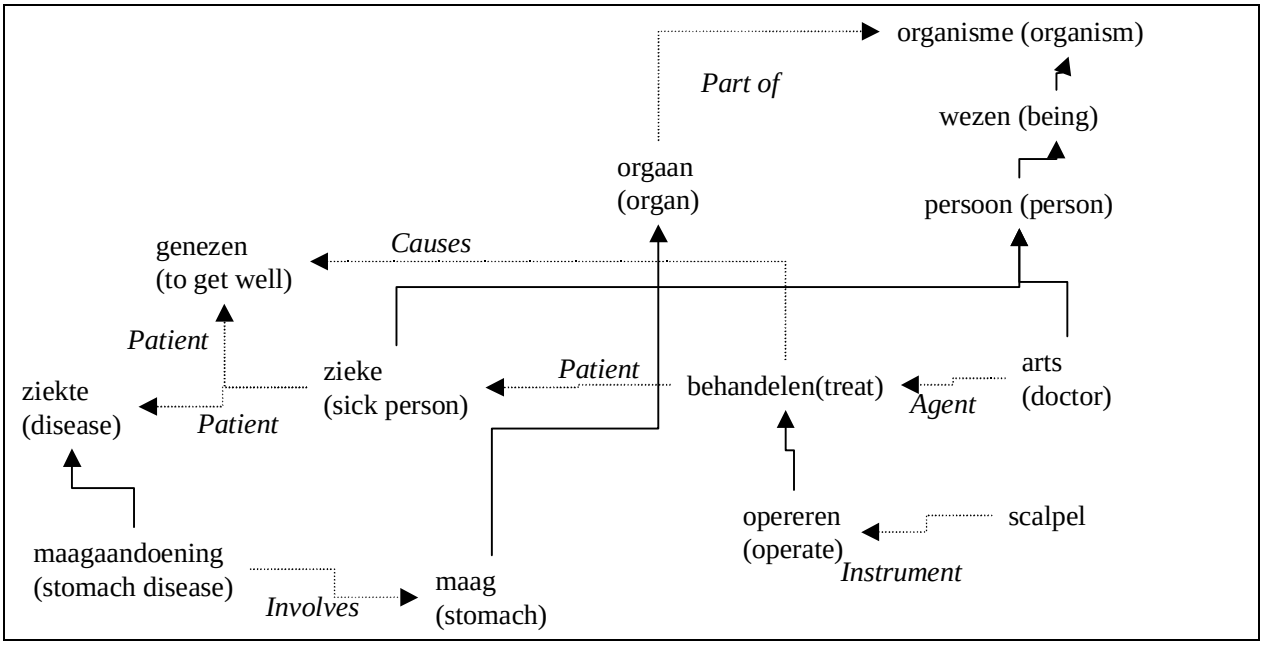


*Figure 7: Thematic network for "disease" in the Dutch wordnet*

Similar graphs can be made for *sports* such as *tennis* and all the involved concepts. By retrieving all concepts related via role/involved and cause relations to an activity, we can thus extract domain information or clusterings from the wordnet internal structure.

Complex causal relations such as the above are not always thematically organized. An important part in the wordnet is related to *communication*, which is perhaps the most complex structure. Figure-8 gives some of the major relations expressed. Communication is an activity where mental content, which is a 3rdOrderEntity, is evoked by a communicative act (2ndOrderEntity), possibly supported by some representation (1stOrderEntity). Basically, a 1stOrderEntity that functions as a symbol contains content as a 3rdOrderEntity (its meaning or interpretation). Communication can use 1stOrder representations as instruments to generate interpretations or meaning. There may be

separate events to create such a representation (e.g. *to utter, to represent, to make an image, to speak, to write*), and there may be communicative events where this representation, created at some other time, is just shown (*to show*). Obviously, information can also be the result of direct perception or by processing representations (e.g. *read, listen, perceive, notice*).

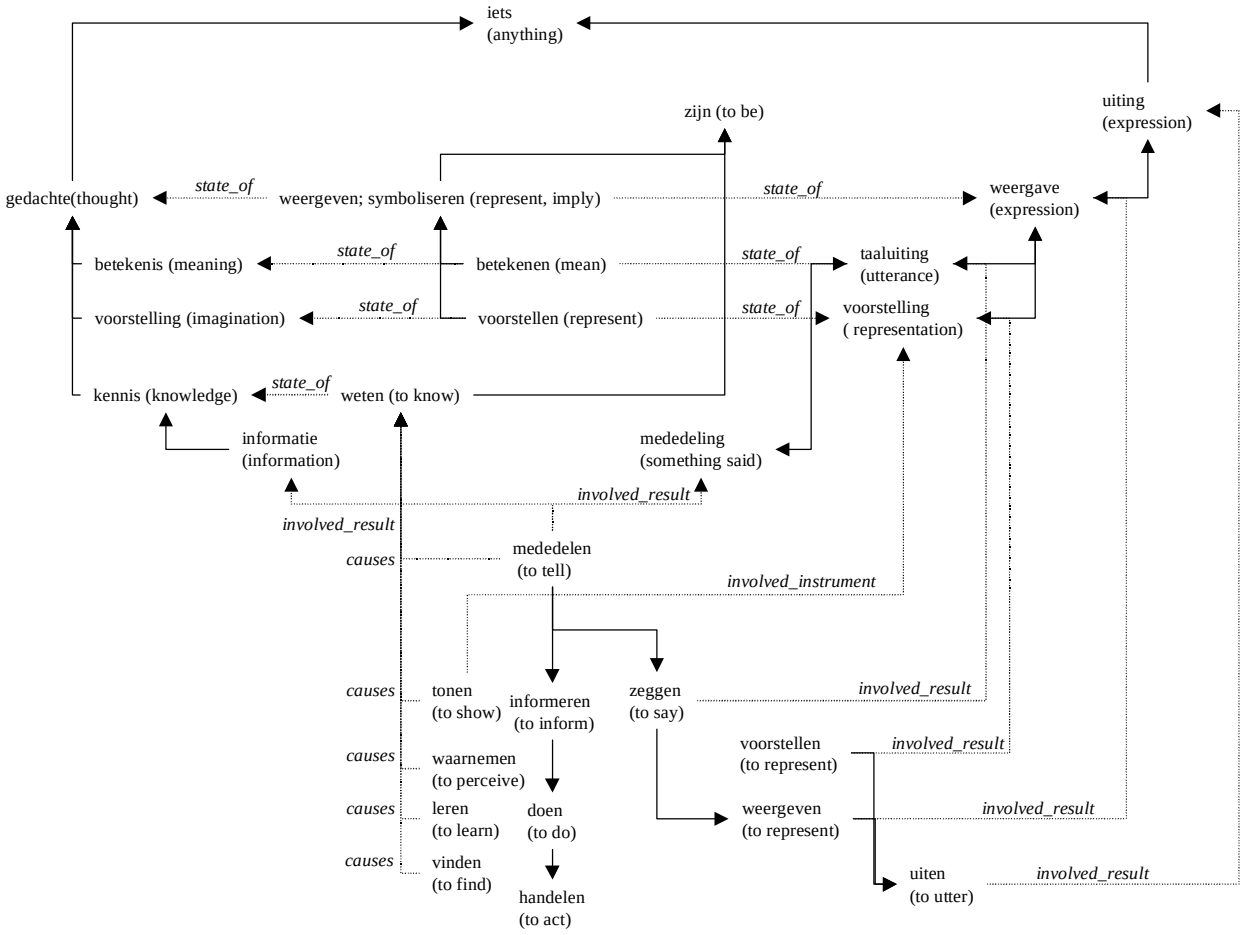


*Figure 8: Causal and involvement relations for "communication"*

At the left side, we see the 3rdOrderEntities below “gedachte” (thought), at the bottom we find some communicative acts as hyponyms of “doen” (to do), some events that cause “weten” (to know), and some events that create a representation, e.g. "weergeven" (to represent), "voorstellen" (to present), "uiten" (express). A complex event is “mededelen” (to tell some information) which results in an utterance (“mededeling”) that informs the addressee (causes the addressee to know (“weten”) and achieve knowledge (“informatie”)). At the right side, we find the representations, which are 1stOrderEntities below “uiting” (expression). Note that "uiting" (expression) can be seen both as a 1st and a 2ndOrderEntity. It can be an event expressing a concept or thought (think of theater performances, signs, body language, etc.) or a symbolic representation with a permanent form. Representations (regardless of their nature) have a relation with their content or meaning (the evoked 3rdOrderEntities). In the middle, we find a set of static verbs (below “zijn” (to be)) that express this relation: “betekenen” (to mean), “weergeven” (to represent), “voorstellen” (to represent). These are transitive verbs where the subject is the symbol and the object is the content.

There are many more relations that have not been represented in this graph, such as: to think ("denken") which may have a *thought* ('gedachte") as a result, *gestures* that are both movements and speech acts but do not result in a 1stOrder representations, *language perceptions* (*listen* and *read*) and *language production* (*write* and *speak*).

### *4.3. Explanation of the EuroWordNet text file*

The synsets in the Dutch wordnet have the general structure of the EuroWordNet format as described in the general EuroWordNet document. There are some specific features that are explained here.

For each variant or synonym in the synset you may find the following information that is specific to the Dutch wordnet:

- Definition in the form of a top-concept classification
- Celex corpus frequency
- Reference to the source of concepts and relations

The so-called central term of a synset in Vlis (most neutral synonym) may have a definition field that contains a classification according to the EuroWordNet top-ontology. These features are derived in the following way:

- The top-ontology features that are assigned to the Base Concepts have been projected to the Dutch representatives of these Base Concepts;
- Other tops and major hyperonyms in the Dutch wordnet have been classified according to the ontology;
- Some large and undifferentiated clusters of concepts have been differentiated using the top-ontology features.

In total, 2006 Dutch synsets received one or more top-concepts (1170 nouns and 836 verbs). These top-concepts have been inherited via the hyponymy relations in the Dutch wordnet, where redundant features have been added. The result is string such as the following:

```
"1stOrderEntity;Artifact;Comestible;Form;Function;Liquid;Living;Natural;
Origin;Part;Substance;"
```

Similar classification has been derived for all the ILI-records using the hyponymy relations of WordNet1.5. However, there are some differences in both the top-concept classification and the hierarchy between the Dutch wordnet and WordNet1.5. This may lead to inconsistencies between the top-concepts in the Dutch definition and the top-concepts in the ILI. Furthermore, the projection of the top-concepts assigned to the Common Base Concepts from the ILI to the Dutch wordnet is not always precise. In some cases, the Common Base Concepts could not be represented by a single synset in the Dutch wordnet but more globally to sets of closely related concepts. For example, there is no equivalent for Base Concept "natural_object" in Dutch, but the hyponyms in WordNet1.5 closely relate to the hyponyms of "voorwerp" (object), which is a bit more general. Linking "natural_object" to "voorwerp" leads to the projection of the feature "Natural", which is then inherited to many thousands of more specific concepts. In some cases,

this over-generates, e.g. other hierarchy nodes below "voorwerp" are marked as Artifact. In these cases, two conflicting top-features are derived.

A explained in section 3, most concepts originate from the Van Dale database Vlis. If this is the case, we have stored a reference to the Vlis sense identifier in the STATUS field of variant. Other information on the source is given as EXTERNAL_INFO in the field SOURCE_ID. The following values occur:

2001 = sense taken from Vlis
2002 = animal or plant sense not in Vlis
2003 = new spelling variant added to an existing concept

Another special feature at EXTERNAL_INFO is the corpus frequency, based on the Celex frequencies. The Celex frequency is derived from a 420K word token corpus of Dutch newspapers. The frequency information is based on stemmed words forms, differentiated per part-of-speech. The type of corpus (Celex) is indicated by the CORPUS_ID number 1 and the actual frequency as the value of FREQUENCY.

Further source information is provided at the relation level (both language internal and equivalence relations), where the following values can occur:

1: language internal relation is copied from Vlis
2: language internal relation is copied from Vlis and checked manually in EuroWordNet
101: language internal relation is added manually in EuroWordNet and checked
102: language internal relation is added manually in EuroWordNet
1001: equivalence relation is added manually and checked
1002: equivalence relation is added manually
1003: equivalence relation is derived automatically and manually checked
1004: equivalence relation is derived automatically and scored as the best two candidates
1005: there are too many automatically derived equuivalences and a reliable parent equivalent relation could be derived.

The appendix shows some entries in the EuroWordNet format for the different parts of speech, in which the above features are illustrated.

## 5. The content of the CD

The Dutch wordnet is delivered as:

- plain text file in the EuroWordNet import/export format: wn_nl.ewn
- Polaris database file: wn_nl.sdb

The Polaris database can be viewed with the freeware program Periscope or with the Polaris database. The Polaris database can be licensed from Lernout and Hauspie (contact person is Geert.Adriaens@lhs.be). The Periscope viewer is on the general CD-rom, as well as instructions how to install the EuroWordNet database and add the Dutch wordnet. Please read the instructions carefully before installing. On the general CD you will also find general documentation on EuroWordNet, and samples of the other wordnets and the complete version of WordNet1.5.

## References

Agirre E. and Rigau G.
1996 Word Sense Disambiguation using Conceptual Density, in proceedings of the 16th International Conference on Computational Linguistics (COLING'96). Copenhagen. 1996.

Boersma, P.
1996 The Sift Lexical Database, Sift LRE 62030, Deliverable D10b, University of Amsterdam. Amsterdam. 1996.

Fellbaum, C. (ed.)
1998 *WordNet: An Electronic Lexical Database*. Cambridge, MA: MIT Press.

Kruyt, T.
1998 "Electronische woordenboeken en tekstcorpora voor Europese taaltechnologie", Trefwoord, 12, 1997-1998, Sdu Uitgevers, Den Haag/ Antwerpen.

Masereeuw, P.
1994 Ltree and CM: The SIFT LDB developer's libraries, Sift deliverable D10, LRE 62030. Computer Centrum Letteren, University of Amsterdam.

Martin W. and J. Tops
1986 Groot woordenboek Engels-Nederlands. Van Dale Lexicografie. Utrecht.

Martin W. and J. Tops
1989 Groot woordenboek Nederlands-Engels. Van Dale Lexicografie. Utrecht.

Miller, G., R. Beckwith, C. Fellbaum, D. Gross and K. Miller.
1990 Five Papers on WordNet. CSL Report 43. Cognitive Science Laboratory. Princeton University.

Procter, P. et al. (Eds)
1978 Longman Dictionary of Contemporary English. Longman, Harlow and London. 1987

Peters W.
1998 "Restructured ILI" EuroWordNet (LE 4003), Deliverable 2D004, University of Sheffield.

Rosch, E. 1977 *Human Categorisation.* In N. Warren (Ed.) Studies in Cross-Cultural Psychology, Vol. I, pp. 1-49. Academic Press. London.

Rigau G., Atserias J. and Agirre E.
1997 Combining Unsupervised Lexical Knowledge Methods for Word Sense Disambiguation, in proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL'97), pages 48-55. Madrid, Spain, 1997.

Rodriquez, H., S. Climent, P. Vossen, L. Bloksma A. Roventini, F. Bertagna, A. Alonge, W. Peters,
The Top-Down Strategy for Building EuroWordNet: Vocabulary Coverage, Base Concepts and Top Ontology. In: Nancy Ide, Daniel Greenstein, Piek Vossen (eds), Special Issue on EuroWordNet. Computers and the Humanities, Volume 32, Nos. 2-3 1998. 117-152.

Sparck-Jones, K. and P. Willett
1997 Readings in Information Retrieval, Morgan Kaufmann Publishers Inc.

Vossen, P., L. Bloksma, H. Rodriguez, S. Climent, N. Calzolari, A. Roventini, F. Bertagna, A. Alonge, W. Peters.

1997 The EuroWordNet Base Concepts and Top Ontology. EuroWordNet (LE 4003) Deliverable D017D034D036. University of Amsterdam

Vossen, P., L. Bloksma, S. Climent, M.A. Marti, G. Oreggioni, G. Escudero, G. Rigau, H. Rodriquez, C. Peters, A. Roventini, F. Bertagna, A. Alonge, W. Peters.
1998 The Restructured Core wordnets in EuroWordNet: Subset1. EuroWordNet (LE 4003), Deliverable D014D015, University of Amsterdam.

Vossen P. (eds.)
1999 EuroWordNet: a multilingual database with lexical semantic networks for European Languages. Kluwer Academic Publishers, Dordrecht.

Vossen, P., L. Bloksma, S. Climent, M.A. Marti, m. Taule, J. Gonzalo, I. Chugur, F. Verdejo, G. Escudero, G. Rigau, H. Rodriquez, A. Alonge, F. Bertagna.
1998 EuroWordNet Restructured Subset2 for Dutch, Spanish and Italian. EuroWordNet (LE 4003), Deliverable D027D028, University of Amsterdam.

Vossen, P., L. Bloksma, S. Climent, M.A. Marti, m. Taule, J. Gonzalo, I. Chugur, F. Verdejo, G. Escudero, G. Rigau, H. Rodriquez, A. Alonge, F. Bertagna.
1999 Comparison of the final wordnets Dutch, Spanish and Italian. EuroWordNet (LE 4003), Deliverable D029D030, University of Amsterdam.

Vossen, P (ed.)
1999 Final Wordnets for Dutch, Spanish, Italian and English. EuroWordNet (LE 4003), Deliverable D032D033, University of Amsterdam.

Vossen, P, P. Boersma and L. Bloksma,
1999 Tools and resources for the Dutch wordnet. EuroWordNet (LE 4003), Internal Report, University of Amsterdam.

Vossen, P (eds.)
2000 EuroWordNet General Document, EuroWordNet (LE2-4003, LE4-8328) University of Amsterdam.

# Appendix I. Examples of Dutch wordnet synsets in EuroWordNet format

```
0 @31066@ WORD_MEANING
  1 PART_OF_SPEECH "n"
  1 VARIANTS
    2 LITERAL "sterkedrank"
      3 SENSE 1
      3 DEFINITION "1stOrderEntity;Artifact;Comestible;Form;Function;Liquid;Living;Natural;
                    Origin;Part;Substance;"
      3 STATUS 272641
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 0
        4 SOURCE_ID 2001
    2 LITERAL "gedistilleerd"
      3 SENSE 2
      3 STATUS 87829
      3 FEATURES
        4 FEATURE "gender"
          5 FEATURE_VALUE "neutral"
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 38
        4 SOURCE_ID 2001
    2 LITERAL "neurie"
      3 SENSE 1
      3 STATUS 414275
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 0
        4 SOURCE_ID 2001
    2 LITERAL "spiritualiën"
      3 SENSE 1
      3 STATUS 281946
      3 FEATURES
        4 FEATURE "number"
          5 FEATURE_VALUE "plural"
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 11
        4 SOURCE_ID 2001
    2 LITERAL "spiritualia"
      3 SENSE 1
      3 STATUS 267141
      3 FEATURES
        4 FEATURE "number"
          5 FEATURE_VALUE "plural"
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 0
        4 SOURCE_ID 2001
    2 LITERAL "vuurwater"
      3 SENSE 1
      3 STATUS 324638
      3 FEATURES
        4 FEATURE "gender"
          5 FEATURE_VALUE "neutral"
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 12
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "has_hyperonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "drank"
          5 SENSE 3
      3 SOURCE_ID 2
    2 RELATION "has_hyponym"
      3 TARGET_CONCEPT
```

```
    4 PART_OF_SPEECH "n"
    4 LITERAL "tic"
      5 SENSE 3
  3 SOURCE_ID 2
2 RELATION "has_hyponym"
  3 TARGET_CONCEPT
    4 PART_OF_SPEECH "n"
    4 LITERAL "mescal"
      5 SENSE 1
  3 SOURCE_ID 2
2 RELATION "has_hyponym"
  3 TARGET_CONCEPT
    4 PART_OF_SPEECH "n"
    4 LITERAL "tequila"
      5 SENSE 1
  3 SOURCE_ID 2
2 RELATION "has_hyponym"
  3 TARGET_CONCEPT
    4 PART_OF_SPEECH "n"
    4 LITERAL "wodka"
      5 SENSE 1
  3 SOURCE_ID 2
2 RELATION "has_hyponym"
  3 TARGET_CONCEPT
    4 PART_OF_SPEECH "n"
    4 LITERAL "aquaviet"
      5 SENSE 1
  3 SOURCE_ID 2
2 RELATION "has_hyponym"
  3 TARGET_CONCEPT
    4 PART_OF_SPEECH "n"
    4 LITERAL "calvados"
      5 SENSE 1
  3 SOURCE_ID 2
2 RELATION "has_hyponym"
  3 TARGET_CONCEPT
    4 PART_OF_SPEECH "n"
    4 LITERAL "pastis"
      5 SENSE 1
  3 SOURCE_ID 2
2 RELATION "has_hyponym"
  3 TARGET_CONCEPT
    4 PART_OF_SPEECH "n"
    4 LITERAL "vieux"
      5 SENSE 1
  3 SOURCE_ID 2
2 RELATION "has_hyponym"
  3 TARGET_CONCEPT
    4 PART_OF_SPEECH "n"
    4 LITERAL "jenever"
      5 SENSE 1
  3 SOURCE_ID 2
2 RELATION "has_hyponym"
  3 TARGET_CONCEPT
    4 PART_OF_SPEECH "n"
    4 LITERAL "likeur"
      5 SENSE 1
  3 SOURCE_ID 1
2 RELATION "has_hyponym"
  3 TARGET_CONCEPT
    4 PART_OF_SPEECH "n"
    4 LITERAL "rum"
      5 SENSE 1
  3 SOURCE_ID 2
2 RELATION "has_hyponym"
  3 TARGET_CONCEPT
    4 PART_OF_SPEECH "n"
    4 LITERAL "slivovitsj"
      5 SENSE 1
  3 SOURCE_ID 2
2 RELATION "has_hyponym"
```

```
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "brandewijn"
          5 SENSE 1
      3 SOURCE_ID 2
    2 RELATION "has_hyponym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "whisky"
          5 SENSE 1
      3 SOURCE_ID 2
    2 RELATION "has_hyponym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "cognac"
          5 SENSE 1
      3 SOURCE_ID 2
  1 EQ_LINKS
    2 EQ_RELATION "eq_synonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "n"
        4 WORDNET_OFFSET 5089006
      3 SOURCE_ID 101

0 @31067@ WORD_MEANING
  1 PART_OF_SPEECH "n"
  1 VARIANTS
    2 LITERAL "sternalgie"
      3 SENSE 1
      3 DEFINITION "2ndOrderEntity;BoundedEvent;Cause;Condition;Dynamic;Experience;
                    Mental;Phenomenal;Physical;SituationType;Static;Stimulating;"
      3 STATUS 272679
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 0
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "has_hyperonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "pijn"
          5 SENSE 1
      3 SOURCE_ID 1
    2 RELATION "involved_location"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "borstbeen"
          5 SENSE 1
      3 SOURCE_ID 101
  1 EQ_LINKS
    2 EQ_RELATION "eq_has_hyperonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "n"
        4 WORDNET_OFFSET 3897733
      3 SOURCE_ID 1005
    2 EQ_RELATION "eq_has_hyperonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "n"
        4 WORDNET_OFFSET 8677522
      3 SOURCE_ID 1005
    2 EQ_RELATION "eq_has_hyperonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "n"
        4 WORDNET_OFFSET 8677932
      3 SOURCE_ID 1005
    2 EQ_RELATION "eq_generalization"
      3 TARGET_ILI
        4 PART_OF_SPEECH "n"
        4 ADD_ON_ID 8559

0 @31068@ WORD_MEANING
```

```
  1 PART_OF_SPEECH "n"
  1 VARIANTS
    2 LITERAL "sterrenkunde"
      3 SENSE 1
      3 DEFINITION "2ndOrderEntity;Agentive;BoundedEvent;Cause;Dynamic;Mental;
                    Purpose;SituationType;Social;UnboundedEvent;"
      3 STATUS 272711
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 49
        4 SOURCE_ID 2001
    2 LITERAL "astronomie"
      3 SENSE 1
      3 STATUS 14387
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 84
        4 SOURCE_ID 2001
    2 LITERAL "uranologie"
      3 SENSE 1
      3 STATUS 303918
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 0
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "has_hyperonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "wetenschap"
          5 SENSE 1
      3 SOURCE_ID 2
    2 RELATION "has_hyponym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "radioastronomie"
          5 SENSE 1
      3 SOURCE_ID 2
    2 RELATION "has_hyponym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "astrodynamica"
          5 SENSE 1
      3 SOURCE_ID 2
    2 RELATION "has_hyponym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "astrometrie"
          5 SENSE 1
      3 SOURCE_ID 2
    2 RELATION "has_hyponym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "astrofysica"
          5 SENSE 1
      3 SOURCE_ID 2
  1 EQ_LINKS
    2 EQ_RELATION "eq_near_synonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "n"
        4 WORDNET_OFFSET 4067256
      3 SOURCE_ID 1004

0 @31069@ WORD_MEANING
  1 PART_OF_SPEECH "n"
  1 VARIANTS
    2 LITERAL "sterrenwacht"
      3 SENSE 2
      3 DEFINITION "1stOrderEntity;Function;Group;Human;Living;Natural;Origin;"
      3 STATUS 272730
      3 EXTERNAL_INFO
```

```
        4 CORPUS_ID 1
          5 FREQUENCY 28
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "has_hyperonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "organisatie"
          5 SENSE 3
      3 SOURCE_ID 2

0 @31081@ WORD_MEANING
  1 PART_OF_SPEECH "n"
  1 VARIANTS
    2 LITERAL "stikstofverbinding"
      3 SENSE 1
      3 DEFINITION "1stOrderEntity;Form;Function;Natural;Origin;Part;Substance;"
      3 STATUS 283078
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 5
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "has_hyperonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "verbinding"
          5 SENSE 3
      3 SOURCE_ID 1
    2 RELATION "has_hyponym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "stikstofdioxyde"
          5 SENSE 1
      3 SOURCE_ID 101
    2 RELATION "has_hyponym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "stikstofmonoxyde"
          5 SENSE 1
      3 SOURCE_ID 101
  1 EQ_LINKS
    2 EQ_RELATION "eq_has_hyperonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "n"
        4 WORDNET_OFFSET 8849147
      3 SOURCE_ID 1005
    2 EQ_RELATION "eq_has_hyperonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "n"
        4 WORDNET_OFFSET 8907331
      3 SOURCE_ID 1005

0 @31082@ WORD_MEANING
  1 PART_OF_SPEECH "n"
  1 VARIANTS
    2 LITERAL "stilist"
      3 SENSE 2
      3 DEFINITION "1stOrderEntity;Animal;Form;Human;Living;Natural;Object;Origin;"
      3 STATUS 273252
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 27
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "has_hyperonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "ontwerper"
          5 SENSE 1
      3 SOURCE_ID 2
```

```
  1 EQ_LINKS
    2 EQ_RELATION "eq_near_synonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "n"
        4 WORDNET_OFFSET 6379239
      3 SOURCE_ID 1004

0 @31083@ WORD_MEANING
  1 PART_OF_SPEECH "n"
  1 VARIANTS
    2 LITERAL "stilstand"
      3 SENSE 2
      3 DEFINITION "2ndOrderEntity;SituationType;Static;"
      3 STATUS 273307
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 609
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "has_hyperonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "toestand"
          5 SENSE 1
      3 SOURCE_ID 2
    2 RELATION "has_hyponym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "bacteriostase"
          5 SENSE 1
      3 SOURCE_ID 101
    2 RELATION "has_hyponym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "stagnatie"
          5 SENSE 1
      3 SOURCE_ID 2
  1 EQ_LINKS
    2 EQ_RELATION "eq_near_synonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "n"
        4 WORDNET_OFFSET 8571118
      3 SOURCE_ID 1004
    2 EQ_RELATION "eq_near_synonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "n"
        4 WORDNET_OFFSET 8572459
      3 SOURCE_ID 1004

0 @31084@ WORD_MEANING
  1 PART_OF_SPEECH "n"
  1 VARIANTS
    2 LITERAL "stilton"
      3 SENSE 1
      3 DEFINITION "1stOrderEntity;Artifact;Comestible;Form;Function;Origin;Solid;Substance;"
      3 STATUS 273322
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 36
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "has_hyperonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "schimmelkaas"
          5 SENSE 1
      3 SOURCE_ID 2
  1 EQ_LINKS
    2 EQ_RELATION "eq_near_synonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "n"
```

```
        4 WORDNET_OFFSET 5052045
      3 SOURCE_ID 1004

0 @2@ WORD_MEANING
  1 PART_OF_SPEECH "v"
  1 VARIANTS
    2 LITERAL "aanaarden"
      3 SENSE 1
      3 DEFINITION "2ndOrderEntity;BoundedEvent;Cause;Dynamic;Experience;Physical;SituationType;"
      3 STATUS 91
      3 FEATURES
        4 FEATURE "trans"
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 0
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "has_hyperonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "v"
        4 LITERAL "bedekken"
          5 SENSE 1
      3 SOURCE_ID 1
    2 RELATION "involved_instrument"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "grond"
          5 SENSE 6
      3 SOURCE_ID 101
  1 EQ_LINKS
    2 EQ_RELATION "eq_near_synonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "v"
        4 WORDNET_OFFSET 867822
      3 SOURCE_ID 1003

0 @3@ WORD_MEANING
  1 PART_OF_SPEECH "v"
  1 VARIANTS
    2 LITERAL "aanbellen"
      3 SENSE 1
      3 DEFINITION "2ndOrderEntity;BoundedEvent;Dynamic;Location;Physical;SituationType;"
      3 STATUS 130
      3 FEATURES
        4 FEATURE "intrans"
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 231
        4 SOURCE_ID 2001
    2 LITERAL "aanschellen"
      3 SENSE 1
      3 STATUS 400542
      3 FEATURES
        4 FEATURE "intrans"
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 2
        4 SOURCE_ID 2001
    2 LITERAL "bellen"
      3 SENSE 1
      3 STATUS 25537
      3 FEATURES
        4 FEATURE "intrans"
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 4292
        4 SOURCE_ID 2001
    2 LITERAL "schellen"
      3 SENSE 1
      3 STATUS 252219
      3 FEATURES
```

```
        4 FEATURE "intrans"
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 29
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "has_hyperonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "v"
        4 LITERAL "melden"
          5 SENSE 2
      3 SOURCE_ID 2
    2 RELATION "involved_instrument"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "deurbel"
          5 SENSE 1
      3 SOURCE_ID 102
  1 EQ_LINKS
    2 EQ_RELATION "eq_near_synonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "v"
        4 WORDNET_OFFSET 1244553
      3 SOURCE_ID 1004
    2 EQ_RELATION "eq_generalization"
      3 TARGET_ILI
        4 PART_OF_SPEECH "v"
        4 ADD_ON_ID 3202
    2 EQ_RELATION "eq_generalization"
      3 TARGET_ILI
        4 PART_OF_SPEECH "v"
        4 ADD_ON_ID 2572

0 @6@ WORD_MEANING
  1 PART_OF_SPEECH "v"
  1 VARIANTS
    2 LITERAL "aanbidden"
      3 SENSE 1
      3 DEFINITION "2ndOrderEntity;Dynamic;Experience;Mental;SituationType;"
      3 STATUS 179
      3 FEATURES
        4 FEATURE "trans"
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 372
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "has_hyperonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "v"
        4 LITERAL "vereren"
          5 SENSE 1
      3 SOURCE_ID 1
    2 RELATION "involved_agent"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "aanbidder"
          5 SENSE 1
      3 SOURCE_ID 101
  1 EQ_LINKS
    2 EQ_RELATION "eq_near_synonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "v"
        4 WORDNET_OFFSET 1012890
      3 SOURCE_ID 1004
    2 EQ_RELATION "eq_near_synonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "v"
        4 WORDNET_OFFSET 1013171
      3 SOURCE_ID 1004
    2 EQ_RELATION "eq_generalization"
```

```
      3 TARGET_ILI
        4 PART_OF_SPEECH "v"
        4 ADD_ON_ID 5066
    2 EQ_RELATION "eq_generalization"
      3 TARGET_ILI
        4 PART_OF_SPEECH "v"
        4 ADD_ON_ID 5080
    2 EQ_RELATION "eq_generalization"
      3 TARGET_ILI
        4 PART_OF_SPEECH "v"
        4 ADD_ON_ID 5206

0 @5@ WORD_MEANING
  1 PART_OF_SPEECH "v"
  1 VARIANTS
    2 LITERAL "aanbetalen"
      3 SENSE 1
      3 DEFINITION "2ndOrderEntity;Agentive;BoundedEvent;Cause;Dynamic;
                    Experience;Location;Physical;Possession;Purpose;SituationType;Social;"
      3 STATUS 150
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 0
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "xpos_near_synonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "aanbetaling"
          5 SENSE 1
      3 SOURCE_ID 101
    2 RELATION "has_hyperonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "v"
        4 LITERAL "betalen"
          5 SENSE 2
      3 SOURCE_ID 1
  1 EQ_LINKS
    2 EQ_RELATION "eq_has_hyperonym"
      3 TARGET_ILI
        4 PART_OF_SPEECH "v"
        4 WORDNET_OFFSET 1281885
      3 SOURCE_ID 1005

0 @9639@ WORD_MEANING
  1 PART_OF_SPEECH "a"
  1 VARIANTS
    2 LITERAL "zwanger"
      3 SENSE 1
      3 STATUS 347356
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 800
        4 SOURCE_ID 2001
    2 LITERAL "gravida"
      3 SENSE 1
      3 STATUS 407647
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 0
        4 SOURCE_ID 2001
    2 LITERAL "pregnant"
      3 SENSE 2
      3 STATUS 223823
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 83
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "xpos_near_synonym"
      3 TARGET_CONCEPT
```

```
        4 PART_OF_SPEECH "n"
        4 LITERAL "zwangerschap"
          5 SENSE 1
      3 SOURCE_ID 101
    2 RELATION "is_caused_by"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "v"
        4 LITERAL "aborteren"
          5 SENSE 1
      3 FEATURES
        4 REVERSED
      3 SOURCE_ID 101
    2 RELATION "is_caused_by"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "v"
        4 LITERAL "bevruchten"
          5 SENSE 1
      3 SOURCE_ID 101

0 @9641@ WORD_MEANING
  1 PART_OF_SPEECH "a"
  1 VARIANTS
    2 LITERAL "insektivoor"
      3 SENSE 2
      3 STATUS 122691
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 0
        4 SOURCE_ID 2001
    2 LITERAL "insectenetend"
      3 SENSE 1
      3 STATUS 0
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 0
        4 SOURCE_ID 2003
    2 LITERAL "insektenetend"
      3 SENSE 1
      3 STATUS 122673
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 5
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "state_of"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "n"
        4 LITERAL "insekteneter"
          5 SENSE 1
      3 FEATURES
        4 REVERSED
      3 SOURCE_ID 102

0 @9642@ WORD_MEANING
  1 PART_OF_SPEECH "a"
  1 VARIANTS
    2 LITERAL "overeenkomend"
      3 SENSE 2
      3 STATUS 208916
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 34
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "xpos_near_synonym"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "v"
        4 LITERAL "lijken"
          5 SENSE 1
      3 SOURCE_ID 101
    2 RELATION "xpos_near_synonym"
```

```
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "v"
        4 LITERAL "overeenkomen"
          5 SENSE 1
      3 SOURCE_ID 101

0 @9595@ WORD_MEANING
  1 PART_OF_SPEECH "b"
  1 VARIANTS
    2 LITERAL "vooruit"
      3 SENSE 1
      3 STATUS 321957
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 2238
        4 SOURCE_ID 2001
    2 LITERAL "voort"
      3 SENSE 1
      3 STATUS 321718
      3 EXTERNAL_INFO
        4 CORPUS_ID 1
          5 FREQUENCY 3987
        4 SOURCE_ID 2001
  1 INTERNAL_LINKS
    2 RELATION "is_caused_by"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "v"
        4 LITERAL "voortgaan"
          5 SENSE 1
      3 FEATURES
        4 REVERSED
      3 SOURCE_ID 101
    2 RELATION "is_caused_by"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "v"
        4 LITERAL "vooruitgaan"
          5 SENSE 3
      3 FEATURES
        4 REVERSED
      3 SOURCE_ID 101
    2 RELATION "is_caused_by"
      3 TARGET_CONCEPT
        4 PART_OF_SPEECH "v"
        4 LITERAL "vooruitzetten"
          5 SENSE 1
      3 FEATURES
        4 REVERSED
      3 SOURCE_ID 101
```

# Appendix II. Major hyperonyms at level 2 in the Dutch wordnet

```
_ iets 1
   HAS_HYPONYM deel 2
        HAS_HYPONYM aandeel 3 NODE 7 / 18
        HAS_HYPONYM afdeling 1 NODE 74 / 302
        HAS_HYPONYM bestanddeel 1 NODE 18 / 30
        HAS_HYPONYM binnenste 1 NODE 10 / 24
        HAS_HYPONYM blad 3 NODE 29 / 153
        HAS_HYPONYM blindedarm 1 NODE 1 / 313
        HAS_HYPONYM bloeiwijze 1 NODE 15 / 21
        HAS_HYPONYM boek 3 NODE 1 / 64
        HAS_HYPONYM boog 2 NODE 13 / 15
        HAS_HYPONYM bot 2 NODE 50 / 79
        HAS_HYPONYM bovenstuk 1 NODE 11 / 30
        HAS_HYPONYM deeltje 1 NODE 16 / 63
        HAS_HYPONYM fundament 1 NODE 10 / 10
        HAS_HYPONYM gebied 2 NODE 91 / 686
        HAS_HYPONYM geheugen 2 NODE 6 / 11
        HAS_HYPONYM gewricht 1 NODE 13 / 13
        HAS_HYPONYM haar 1 NODE 18 / 53
        HAS_HYPONYM haar 2 NODE 15 / 18
        HAS_HYPONYM inrichting 5 NODE 30 / 79
        HAS_HYPONYM insnijding 1 NODE 15 / 25
        HAS_HYPONYM kant 1 NODE 27 / 52
        HAS_HYPONYM knol 1 NODE 7 / 33
        HAS_HYPONYM laag 1 NODE 73 / 170
        HAS_HYPONYM lichaamsdeel 1 NODE 48 / 213
        HAS_HYPONYM onderdeel 1 NODE 53 / 250
        HAS_HYPONYM opening 3 NODE 32 / 301
        HAS_HYPONYM orgaan 1 NODE 44 / 312
        HAS_HYPONYM overblijfsel 2 NODE 9 / 54
        HAS_HYPONYM overschot 2 NODE 11 / 11
        HAS_HYPONYM passage 1 NODE 11 / 15
        HAS_HYPONYM perk 1 NODE 5 / 15
        HAS_HYPONYM plaat 1 NODE 56 / 237
        HAS_HYPONYM plooi 1 NODE 4 / 22
        HAS_HYPONYM post 2 NODE 12 / 12
        HAS_HYPONYM rand 1 NODE 41 / 116
        HAS_HYPONYM rang 2 NODE 9 / 19
        HAS_HYPONYM stengel 1 NODE 9 / 16
        HAS_HYPONYM steun 1 NODE 28 / 72
        HAS_HYPONYM strook 1 NODE 26 / 309
        HAS_HYPONYM tak 1 NODE 15 / 29
        HAS_HYPONYM tand 1 NODE 11 / 22
        HAS_HYPONYM uiteinde 1 NODE 28 / 59
        HAS_HYPONYM verbinding 5 NODE 20 / 63
        HAS_HYPONYM vezel 1 NODE 12 / 18
        HAS_HYPONYM vlak 1 NODE 36 / 233
        HAS_HYPONYM vrucht 1 NODE 15 / 194
        HAS_HYPONYM water 2 NODE 16 / 106
        HAS_HYPONYM wijk 1 NODE 21 / 26
        HAS_HYPONYM wortel 1 NODE 10 / 17
   HAS_HYPONYM eigenschap 1
        HAS_HYPONYM aard 1 NODE 6 / 10
        HAS_HYPONYM karaktertrek 1 NODE 58 / 218
        HAS_HYPONYM kleur 1 NODE 31 / 63
        HAS_HYPONYM kwaliteit 1 NODE 12 / 20
        HAS_HYPONYM maat 5 NODE 27 / 1544
        HAS_HYPONYM niveau 1 NODE 24 / 57
        HAS_HYPONYM onduidelijkheid 1 NODE 6 / 11
        HAS_HYPONYM onvolkomenheid 1 NODE 10 / 684
        HAS_HYPONYM overeenkomst 1 NODE 8 / 44
        HAS_HYPONYM schoonheid 1 NODE 3 / 11
        HAS_HYPONYM stand 2 NODE 3 / 69
        HAS_HYPONYM stijfheid 1 NODE 8 / 11
        HAS_HYPONYM structuur 2 NODE 25 / 271
        HAS_HYPONYM vermogen 3 NODE 66 / 331
        HAS_HYPONYM vorm 1 NODE 36 / 126
```

```
     HAS_HYPONYM waarde 1 NODE 29 / 35
     HAS_HYPONYM waarde 2 NODE 9 / 18
     HAS_HYPONYM wijze 2 NODE 49 / 420
HAS_HYPONYM gebeurtenis 1
     HAS_HYPONYM bevalling 1 NODE 9 / 10
     HAS_HYPONYM bezigheid 1 NODE 58 / 3191
     HAS_HYPONYM ervaring 2 NODE 10 / 531
     HAS_HYPONYM feit 1 NODE 8 / 69
     HAS_HYPONYM geval 2 NODE 11 / 11
     HAS_HYPONYM handeling 1 NODE 90 / 4041
     HAS_HYPONYM ongeluk 2 NODE 10 / 32
     HAS_HYPONYM rechtszaak 1 NODE 13 / 13
     HAS_HYPONYM verandering 1 NODE 127 / 3782
     HAS_HYPONYM verkiezingen 1 NODE 10 / 10
     HAS_HYPONYM verschijnsel 1 NODE 22 / 1140
     HAS_HYPONYM vervolg 1 NODE 3 / 12
     HAS_HYPONYM voorval 1 NODE 4 / 14
     HAS_HYPONYM wedstrijd 1 NODE 84 / 208
     HAS_HYPONYM wedstrijdonderdeel 1 NODE 12 / 27
HAS_HYPONYM gedachte 2
     HAS_HYPONYM betekenis 1 NODE 12 / 68
     HAS_HYPONYM doel 3 NODE 8 / 15
     HAS_HYPONYM idee 1 NODE 15 / 197
     HAS_HYPONYM kennis 2 NODE 27 / 687
     HAS_HYPONYM mening 1 NODE 16 / 171
     HAS_HYPONYM overpeinzing 1 NODE 5 / 6
     HAS_HYPONYM reden 1 NODE 6 / 79
     HAS_HYPONYM redenering 1 NODE 9 / 11
     HAS_HYPONYM vertrouwen 1 NODE 7 / 19
     HAS_HYPONYM voorstelling 1 NODE 26 / 57
HAS_HYPONYM geheel 1
     HAS_HYPONYM complex 1 NODE 6 / 15
     HAS_HYPONYM leer 3 NODE 16 / 69
     HAS_HYPONYM opeenhoping 2 NODE 7 / 8
     HAS_HYPONYM reglement 1 NODE 14 / 14
     HAS_HYPONYM samenstel 1 NODE 14 / 30
     HAS_HYPONYM stelsel 1 NODE 20 / 69
     HAS_HYPONYM techniek 2 NODE 14 / 14
     HAS_HYPONYM verzameling 2 NODE 60 / 348
HAS_HYPONYM groep 4
     HAS_HYPONYM afdeling 1 NODE 74 / 302
     HAS_HYPONYM apparaat 2 NODE 9 / 10
     HAS_HYPONYM apparatuur 1 NODE 20 / 30
     HAS_HYPONYM bezit 1 NODE 24 / 57
     HAS_HYPONYM eetgerei 1 NODE 8 / 150
     HAS_HYPONYM goed 4 NODE 16 / 699
     HAS_HYPONYM goederen 2 NODE 23 / 204
     HAS_HYPONYM groen 3 NODE 4 / 14
     HAS_HYPONYM groep 3 NODE 140 / 1176
     HAS_HYPONYM handelswaar 1 NODE 14 / 27
     HAS_HYPONYM installatie 3 NODE 39 / 264
     HAS_HYPONYM koppel 2 NODE 2 / 10
     HAS_HYPONYM maatschappij 1 NODE 14 / 14
     HAS_HYPONYM materiaal 3 NODE 22 / 448
     HAS_HYPONYM organisatie 3 NODE 79 / 1226
     HAS_HYPONYM post 7 NODE 14 / 15
     HAS_HYPONYM reeks 1 NODE 18 / 261
     HAS_HYPONYM sector 3 NODE 43 / 107
     HAS_HYPONYM set 2 NODE 7 / 22
     HAS_HYPONYM software 1 NODE 7 / 22
     HAS_HYPONYM speelgoed 1 NODE 35 / 63
     HAS_HYPONYM stel 1 NODE 8 / 47
     HAS_HYPONYM stelsel 1 NODE 20 / 69
     HAS_HYPONYM toebehoren 1 NODE 4 / 50
     HAS_HYPONYM vee 1 NODE 13 / 16
     HAS_HYPONYM verzameling 2 NODE 60 / 348
     HAS_HYPONYM wereld 2 NODE 11 / 20
     HAS_HYPONYM woordgroep 1 NODE 5 / 50
HAS_HYPONYM hoeveelheid 1
     HAS_HYPONYM aandeel 3 NODE 7 / 18
     HAS_HYPONYM beetje 1 NODE 7 / 12
```

```
      HAS_HYPONYM bos 1 NODE 5 / 12
      HAS_HYPONYM brok 2 NODE 1 / 68
      HAS_HYPONYM bundel 1 NODE 18 / 39
      HAS_HYPONYM deeltje 2 NODE 2 / 10
      HAS_HYPONYM druppel 1 NODE 11 / 12
      HAS_HYPONYM hoop 1 NODE 11 / 25
      HAS_HYPONYM laag 1 NODE 73 / 170
      HAS_HYPONYM monster 1 NODE 8 / 10
      HAS_HYPONYM opbrengst 1 NODE 18 / 45
      HAS_HYPONYM percentage 1 NODE 13 / 17
      HAS_HYPONYM plak 1 NODE 5 / 18
      HAS_HYPONYM portie 2 NODE 10 / 49
      HAS_HYPONYM produktie 2 NODE 11 / 11
      HAS_HYPONYM rol 2 NODE 10 / 10
      HAS_HYPONYM voorraad 1 NODE 9 / 23
      HAS_HYPONYM zaad 2 NODE 18 / 21
   HAS_HYPONYM middel 2
      HAS_HYPONYM apparaat 1 NODE 239 / 842
      HAS_HYPONYM bedekking 1 NODE 24 / 574
      HAS_HYPONYM betaalmiddel 1 NODE 9 / 99
      HAS_HYPONYM communicatiemiddel 1 NODE 3 / 177
      HAS_HYPONYM computerprogramma 1 NODE 13 / 14
      HAS_HYPONYM hulpmiddel 1 NODE 19 / 57
      HAS_HYPONYM informatiedrager 1 NODE 7 / 50
      HAS_HYPONYM instrument 1 NODE 39 / 604
      HAS_HYPONYM kleefstof 1 NODE 4 / 23
      HAS_HYPONYM kleurstof 1 NODE 22 / 66
      HAS_HYPONYM lokaas 1 NODE 7 / 11
      HAS_HYPONYM materiaal 3 NODE 22 / 448
      HAS_HYPONYM middel 3 NODE 82 / 580
      HAS_HYPONYM software 1 NODE 7 / 22
      HAS_HYPONYM vervoermiddel 1 NODE 8 / 692
      HAS_HYPONYM weg 2 NODE 5 / 120
      HAS_HYPONYM werktuig 1 NODE 109 / 810
   HAS_HYPONYM object 1
      HAS_HYPONYM bouwwerk 1 NODE 12 / 575
      HAS_HYPONYM hemellichaam 1 NODE 7 / 40
      HAS_HYPONYM voorwerp 1 NODE 193 / 8474
      HAS_HYPONYM wezen 1 NODE 40 / 8893
   HAS_HYPONYM plaats 1
      HAS_HYPONYM aanlegplaats 1 NODE 5 / 33
      HAS_HYPONYM adres 1 NODE 9 / 11
      HAS_HYPONYM baan 3 NODE 24 / 32
      HAS_HYPONYM basis 5 NODE 8 / 11
      HAS_HYPONYM gebied 2 NODE 91 / 686
      HAS_HYPONYM gelegenheid 2 NODE 29 / 121
      HAS_HYPONYM laagte 1 NODE 2 / 23
      HAS_HYPONYM land 1 NODE 16 / 51
      HAS_HYPONYM plaats 5 NODE 7 / 15
      HAS_HYPONYM plek 1 NODE 18 / 62
      HAS_HYPONYM post 3 NODE 12 / 18
      HAS_HYPONYM punt 6 NODE 35 / 97
      HAS_HYPONYM ruimte 1 NODE 14 / 151
      HAS_HYPONYM ruimte 3 NODE 93 / 1457
      HAS_HYPONYM station 1 NODE 10 / 10
      HAS_HYPONYM verblijfplaats 1 NODE 32 / 721
      HAS_HYPONYM verhevenheid 2 NODE 11 / 116
      HAS_HYPONYM water 2 NODE 16 / 106
      HAS_HYPONYM weg 1 NODE 56 / 158
   HAS_HYPONYM produkt 1
      HAS_HYPONYM artikel 4 NODE 27 / 157
      HAS_HYPONYM bouwwerk 1 NODE 12 / 575
      HAS_HYPONYM computerprogramma 1 NODE 13 / 14
      HAS_HYPONYM constructie 3 NODE 19 / 179
      HAS_HYPONYM distillaat 1 NODE 2 / 16
      HAS_HYPONYM ei 2 NODE 10 / 11
      HAS_HYPONYM handwerkje 1 NODE 6 / 15
      HAS_HYPONYM mengsel 1 NODE 52 / 230
      HAS_HYPONYM natuurprodukt 1 NODE 6 / 48
      HAS_HYPONYM software 1 NODE 7 / 22
      HAS_HYPONYM teelt 2 NODE 2 / 53
```

```
     HAS_HYPONYM weg 1 NODE 56 / 158
     HAS_HYPONYM werk 4 NODE 16 / 829
     HAS_HYPONYM zuivelprodukt 1 NODE 10 / 91
HAS_HYPONYM soort 2
     HAS_HYPONYM diersoort 1 NODE 8 / 13
     HAS_HYPONYM gewas 1 NODE 22 / 96
     HAS_HYPONYM graansoort 1 NODE 5 / 15
     HAS_HYPONYM houtsoort 1 NODE 56 / 57
HAS_HYPONYM substantie 1
     HAS_HYPONYM aanslag 5 NODE 8 / 14
     HAS_HYPONYM afscheiding 3 NODE 16 / 63
     HAS_HYPONYM brandstof 1 NODE 18 / 67
     HAS_HYPONYM deeltje 1 NODE 16 / 63
     HAS_HYPONYM delfstof 1 NODE 12 / 55
     HAS_HYPONYM element 1 NODE 4 / 115
     HAS_HYPONYM element 2 NODE 96 / 176
     HAS_HYPONYM gas 1 NODE 31 / 71
     HAS_HYPONYM grond 6 NODE 29 / 83
     HAS_HYPONYM hars 1 NODE 11 / 16
     HAS_HYPONYM hormoon 1 NODE 17 / 20
     HAS_HYPONYM kleefstof 1 NODE 4 / 23
     HAS_HYPONYM kleurstof 1 NODE 22 / 66
     HAS_HYPONYM massa 3 NODE 12 / 62
     HAS_HYPONYM materiaal 1 NODE 35 / 843
     HAS_HYPONYM mengsel 1 NODE 52 / 230
     HAS_HYPONYM mineraal 1 NODE 5 / 10
     HAS_HYPONYM neerslag 1 NODE 10 / 18
     HAS_HYPONYM papier 1 NODE 64 / 96
     HAS_HYPONYM poeder 2 NODE 20 / 24
     HAS_HYPONYM verbinding 3 NODE 82 / 440
     HAS_HYPONYM vloeistof 1 NODE 42 / 555
     HAS_HYPONYM voedsel 1 NODE 66 / 1327
     HAS_HYPONYM vulling 1 NODE 13 / 16
     HAS_HYPONYM weefsel 2 NODE 34 / 399
HAS_HYPONYM tijd 1
     HAS_HYPONYM periode 1 NODE 129 / 419
     HAS_HYPONYM tijd 3 NODE 50 / 70
     HAS_HYPONYM tijdeenheid 1 NODE 20 / 235
     HAS_HYPONYM tijdstip 1 NODE 23 / 81
HAS_HYPONYM toestand 1
     HAS_HYPONYM afwezigheid 1 NODE 11 / 13
     HAS_HYPONYM afwijking 1 NODE 15 / 33
     HAS_HYPONYM behoefte 1 NODE 7 / 10
     HAS_HYPONYM bepaling 2 NODE 14 / 234
     HAS_HYPONYM bezit 2 NODE 10 / 11
     HAS_HYPONYM gesteldheid 1 NODE 22 / 1946
     HAS_HYPONYM grootte 1 NODE 9 / 55
     HAS_HYPONYM hoedanigheid 1 NODE 6 / 59
     HAS_HYPONYM houding 1 NODE 31 / 63
     HAS_HYPONYM moeilijkheid 1 NODE 3 / 28
     HAS_HYPONYM mogelijkheid 1 NODE 29 / 102
     HAS_HYPONYM niveau 1 NODE 24 / 57
     HAS_HYPONYM omstandigheid 1 NODE 18 / 182
     HAS_HYPONYM ongemak 1 NODE 1 / 15
     HAS_HYPONYM overeenkomst 1 NODE 8 / 44
     HAS_HYPONYM overeenkomst 2 NODE 30 / 200
     HAS_HYPONYM plicht 1 NODE 29 / 46
     HAS_HYPONYM positie 3 NODE 12 / 19
     HAS_HYPONYM relatie 1 NODE 8 / 152
     HAS_HYPONYM resultaat 1 NODE 11 / 85
     HAS_HYPONYM ruimte 1 NODE 14 / 151
     HAS_HYPONYM rust 3 NODE 4 / 16
     HAS_HYPONYM schade 2 NODE 23 / 24
     HAS_HYPONYM toestemming 1 NODE 11 / 85
     HAS_HYPONYM vrijheid 1 NODE 9 / 17
     HAS_HYPONYM wijze 2 NODE 49 / 420
HAS_HYPONYM uiting 2
     HAS_HYPONYM opmerking 1 NODE 26 / 41
     HAS_HYPONYM taaluiting 1 NODE 9 / 1918
     HAS_HYPONYM weergave 2 NODE 10 / 3916
     HAS_HYPONYM wens 2 NODE 2 / 14
```

```
_ gebeuren 2
   HAS_HYPONYM handelen 2
        HAS_HYPONYM aanvallen 2 NODE 14 / 18
        HAS_HYPONYM bezighouden 3 NODE 21 / 444
        HAS_HYPONYM doen 5 NODE 47 / 3126
        HAS_HYPONYM gedragen 3 NODE 32 / 207
        HAS_HYPONYM haasten 1 NODE 1 / 24
        HAS_HYPONYM helpen 1 NODE 20 / 113
        HAS_HYPONYM inspannen 3 NODE 5 / 14
        HAS_HYPONYM leiden 4 NODE 12 / 20
        HAS_HYPONYM leren 3 NODE 6 / 36
        HAS_HYPONYM oefenen 3 NODE 6 / 9
        HAS_HYPONYM optreden 4 NODE 7 / 21
        HAS_HYPONYM overtreffen 1 NODE 8 / 18
        HAS_HYPONYM praten 2 NODE 11 / 30
        HAS_HYPONYM proberen 2 NODE 15 / 101
        HAS_HYPONYM reageren 2 NODE 7 / 22
        HAS_HYPONYM vechten 1 NODE 14 / 21
        HAS_HYPONYM vergissen 1 NODE 10 / 10
        HAS_HYPONYM werken 1 NODE 20 / 85
        HAS_HYPONYM zorgen 1 NODE 3 / 18
   HAS_HYPONYM meemaken 1
        HAS_HYPONYM ervaren 2 NODE 8 / 90
        HAS_HYPONYM ondergaan 1 NODE 8 / 20
        HAS_HYPONYM teleurstellen 1 NODE 4 / 5
        HAS_HYPONYM waarnemen 2 NODE 9 / 132
   HAS_HYPONYM uitoefenen 2
        HAS_HYPONYM duwen 1 NODE 9 / 93
        HAS_HYPONYM raken 2 NODE 9 / 232
        HAS_HYPONYM trekken 1 NODE 6 / 77
        HAS_HYPONYM zuigen 2 NODE 5 / 15
   HAS_HYPONYM veranderen 1
        HAS_HYPONYM achteruitgaan 2 NODE 17 / 31
        HAS_HYPONYM bewegen 1 NODE 36 / 1604
        HAS_HYPONYM branden 1 NODE 9 / 16
        HAS_HYPONYM doodgaan 1 NODE 2 / 21
        HAS_HYPONYM kapotgaan 1 NODE 16 / 51
        HAS_HYPONYM krijgen 1 NODE 15 / 25
        HAS_HYPONYM loslaten 1 NODE 19 / 26
        HAS_HYPONYM ontstaan 1 NODE 14 / 28
        HAS_HYPONYM opengaan 1 NODE 13 / 14
        HAS_HYPONYM resulteren 1 NODE 7 / 44
        HAS_HYPONYM toenemen 1 NODE 23 / 66
        HAS_HYPONYM uiteenvallen 1 NODE 10 / 26
        HAS_HYPONYM verbeteren 1 NODE 6 / 47
        HAS_HYPONYM verdwijnen 1 NODE 21 / 35
        HAS_HYPONYM verminderen 1 NODE 11 / 25
        HAS_HYPONYM verschijnen 1 NODE 10 / 19
        HAS_HYPONYM vervormen 1 NODE 7 / 12
   HAS_HYPONYM veroorzaken 1
        HAS_HYPONYM aanvoelen 1 NODE 5 / 12
        HAS_HYPONYM be\e"indigen 1 NODE 36 / 114
        HAS_HYPONYM bereiken 2 NODE 21 / 373
        HAS_HYPONYM bewerkstelligen 1 NODE 2 / 49
        HAS_HYPONYM bezorgen 2 NODE 5 / 19
        HAS_HYPONYM draaien 11 NODE 12 / 21
        HAS_HYPONYM klinken 1 NODE 61 / 88
        HAS_HYPONYM laten 8 NODE 18 / 84
        HAS_HYPONYM maken 2 NODE 46 / 868
        HAS_HYPONYM omvergooien 1 NODE 18 / 21
        HAS_HYPONYM veranderen 4 NODE 124 / 4214
        HAS_HYPONYM verhinderen 1 NODE 7 / 31
        HAS_HYPONYM verwerven 1 NODE 13 / 153
        HAS_HYPONYM voortbrengen 1 NODE 6 / 71
        HAS_HYPONYM vormen 2 NODE 5 / 10
```

```
_ zijn 7
   HAS_HYPONYM bevinden 2
        HAS_HYPONYM bevinden 3 NODE 27 / 154
        HAS_HYPONYM hebben 1 NODE 10 / 41
        HAS_HYPONYM slapen 1 NODE 8 / 10
        HAS_HYPONYM vastzitten 1 NODE 3 / 10
   HAS_HYPONYM denken 4
        HAS_HYPONYM beschouwen 2 NODE 17 / 20
        HAS_HYPONYM betreuren 1 NODE 2 / 3
        HAS_HYPONYM instemmen 1 NODE 5 / 5
        HAS_HYPONYM speculeren 2 NODE 1 / 1
        HAS_HYPONYM vermoeden 2 NODE 1 / 3
        HAS_HYPONYM veronderstellen 1 NODE 2 / 2
        HAS_HYPONYM vertrouwen 2 NODE 5 / 10
   HAS_HYPONYM eruitzien 1
        HAS_HYPONYM golven 2 NODE 1 / 2
        HAS_HYPONYM hellen 2 NODE 2 / 4
        HAS_HYPONYM kronkelen 1 NODE 1 / 1
        HAS_HYPONYM omkransen 1 NODE 1 / 1
        HAS_HYPONYM staan 3 NODE 2 / 2
        HAS_HYPONYM stralen 2 NODE 1 / 1
        HAS_HYPONYM uitsteken 1 NODE 7 / 15
   HAS_HYPONYM zijn 2
        HAS_HYPONYM bestaan 3 NODE 2 / 9
        HAS_HYPONYM gelden 3 NODE 2 / 2
        HAS_HYPONYM leven 7 NODE 3 / 4
        HAS_HYPONYM voorkomen 5 NODE 1 / 1
```