

Documentation

Version 7

<http://www.let.vu.nl/onderzoek/projectsites/cornetto/start.htm>

nũ | STEVIN



Table of contents

1	INTRODUCTION	6
2	PROJECT OVERVIEW	7
3	STATISTICS.....	12
3.1	STATISTICS FOR SYNSETS	13
3.2	STATISTICS FOR LEXICAL UNITS.....	18
3.3	STATISTICS FOR CID RECORDS	19
4	DATABASE DESIGN AND ALIGNMENT.....	21
4.1	DESIGN	21
4.2	ALIGNMENT QUALITY LABEL	24
5	DATA COLLECTIONS.....	28
5.1	CORNETTO SYNSETS	28
5.1.1	<i>Nouns</i>	28
5.1.1.1	ID number	28
5.1.1.2	Synonyms.....	29
5.1.1.3	Internal Relations	31
5.1.1.4	Wordnet Equivalence relations.....	33
5.1.1.5	Ontology.....	35
5.1.1.6	Wordnet domains	41
5.1.2	<i>Verbs</i>	42
5.1.2.1	ID Number	42
5.1.2.2	Synonyms.....	43
5.1.2.3	Internal relations.....	43
5.1.2.4	Wordnet Equivalence relations.....	43
5.1.2.5	Ontology.....	43
5.1.2.6	Wordnet domains	43
5.1.3	<i>Adjectives</i>	43
5.1.3.1	ID Number	44
5.1.3.2	Synonyms.....	44
5.1.3.3	Internal relations.....	44
5.1.3.4	Wordnet equivalence relations	44
5.1.3.5	Ontology.....	45
5.1.3.6	Domains	47
5.2	CORNETTO LEXICAL UNITS.....	47
5.2.1	<i>Nouns</i>	48
5.2.1.1	Lexical unit ID, and sequence number	48
5.2.1.2	Morphology.....	49
5.2.1.3	Syntax	49



5.2.1.4	Semantics	50
5.2.1.5	Pragmatics	51
5.2.1.6	Examples	52
5.2.2	<i>Verbs</i>	55
5.2.2.1	Lexical unit ID, and sequence number	55
5.2.2.2	Morphology	56
5.2.2.3	Syntax	56
5.2.2.4	Semantics	58
5.2.2.5	Pragmatics	59
5.2.2.6	Examples	59
5.2.3	<i>Adjectives</i>	59
5.2.3.1	Lexical unit ID, and sequence number	59
5.2.3.2	Morphology	59
5.2.3.3	Syntax	60
5.2.3.4	Semantics	61
5.2.3.5	Pragmatics	62
5.2.3.6	Examples	62
6	REFERENCES	62
7	APPENDICES.....	64
7.1	APPENDIX A: WORDNETDOMAIN HIEARCHY.....	64
7.2	APPENDIX B: MANUALLY ALIGNED WORDS.....	69
7.3	NOUNS WITH 8 OR MORE EQUIVALENCE RELATIONS THAT ARE POST-EDITED MANUALLY FOR EQUIVALENCE, SUMO LABEL AND DOMAIN LABEL	85
7.4	VERBS WITH 4 OR MORE EQUIVALENCE RELATIONS THAT ARE POST-EDITED MANUALLY FOR EQUIVALENCE, SUMO LABEL AND DOMAIN LABEL	87



List of Tables

Table 1: Form units (FUs) in the original source databases.....	9
Table 2: Overview data Cornetto repositories.....	12
Table 3: Overall data fields for synsets.....	13
Table 4: Distribution of synset relations	14
Table 5: Distribution of equivalence relations	14
Table 6: Distribution of Wordnet Domain tags.....	16
Table 7: Miscellaneous data for synsets.....	17
Table 8: Example fields.....	18
Table 9: Pragmatics fields	18
Table 10: Data elements for syntax, semantics and morphology.....	18
Table 11: Number of matching and non-matching Lexical Units.....	19
Table 12: Status values for CID records	20
Table 13: Author labels for CID mappings	20
Table 14: Alignment quality labels	25
Table 15: SUMO relations used in Cornetto	36
Table 16: SUMO terms used for Cornetto Adjectives	45



Consortium:

Vrije Universiteit Amsterdam, Faculteit der Letteren (VU-AMS)

Universiteit van Amsterdam, Instituut voor Informatica (UVA-AMS)

Katholieke Universiteit Leuven-ICRI (KU-LEU)

Irion Technologies B.V. (IRION)

Persons involved:

Piek Vossen (VU-AMS/IRION)

Willy Martin (VU-AMS)

Isa Maks (VU-AMS)

Roxane Segers (VU-AMS)

Hennie van der Vliet (VU-AMS)

Maarten de Rijke (UVA-AMS)

Erik Tjong Kim Sang (UVA-AMS)

Katja Hofmann (UVA-AMS)

Marie-Francine Moens (KU-LEU)

Erik Boiy (KU-LEU)

Hetty van Zutphen (IRION)

Agata Cybulska (IRION)

Anja Weisscher (Vulpes Language Holding BV)



1 Introduction

Cornetto is a lexical semantic database for Dutch, covering over 92K lemmas (70K nouns, 9K verbs, 12K adjectives and 73 adverbs) corresponding to 118K word meanings. The database includes the most generic and central part of the language. The database goes beyond the structure and content of Wordnet and FrameNet. It contains both vertical and horizontal semantic relations and combinatorial lexical constraints such as multiword expressions, idioms and collocations on the one hand, and lexical functions and frames on the other. The concepts are aligned with the English Wordnet so that ontologies and domain labels are also imported. The semantic layer is validated with a formal ontology, to make it usable in Semantic Web environments. The lexical database is evaluated by integration in IR and QA applications (see the Cornetto deliverables D13 and D14).

In addition to the Cornetto database, there is a toolkit for the acquisition of new concepts and relations, and the tuning and extraction of a domain specific sub-lexicon from a compiled corpus. Such a sub-lexicon is extracted for the domain of financial law.

The Cornetto goals fit the resources priority for electronic lexicons and the research priority for Semantic analysis. In the area of applications it is related to:

- Monolingual and multilingual information extraction
- Semantic web
- Dialogue and QA solutions
- Automatic summarization and text generation applications
- Machine translation
- Educational systems

The Cornetto project started January 2006 and the first version was released in September 2008.

This deliverable starts with a project overview and then will mainly focus on the database design, database contents and the data collections.



2 Project Overview

In short, the results of the Cornetto project consist of three major components:

1. The Cornetto database in XML format, including XSD schemas for lexical units, synsets and so-called CID records, and as an online DebVisDic server (deliverable D01)
2. The general acquisition toolkit (deliverables D06, D07 and D08)
3. The domain-acquisition toolkit & acquired legal domain vocabulary (deliverables D09, D10 and D11)

The Cornetto databases contain:

- 92,686 lemmas, including the most generic and central part of the language and a specialized database for the domain of financial law.
- vertical semantic relations, e.g. hyponym and synonym relations;
- horizontal semantic relations, such as roles, part-whole relations, causal relations;
- combinatorial relations, such as lexical functions, selectional restrictions, collocations, syntactic-semantic frames;
- a top-level ontology;
- an ontological typing of the lexical units;
- a domain ontology (Wordnet Domains);
- a domain labelling of the lexical units;
- an equivalence mapping to synsets in WordNet2.0 and WordNet3.0;
- morpho-syntactic information, including syntactic complementation frames;

Furthermore, Cornetto includes:

- An open-source and public database system with editor, import/export functions and API;
- The methodology and toolkit for acquiring new concepts and relations from corpora;
- The methodology and toolkit for tuning and customizing to a specific domain;
- A series of deliverables describing the work and the results.

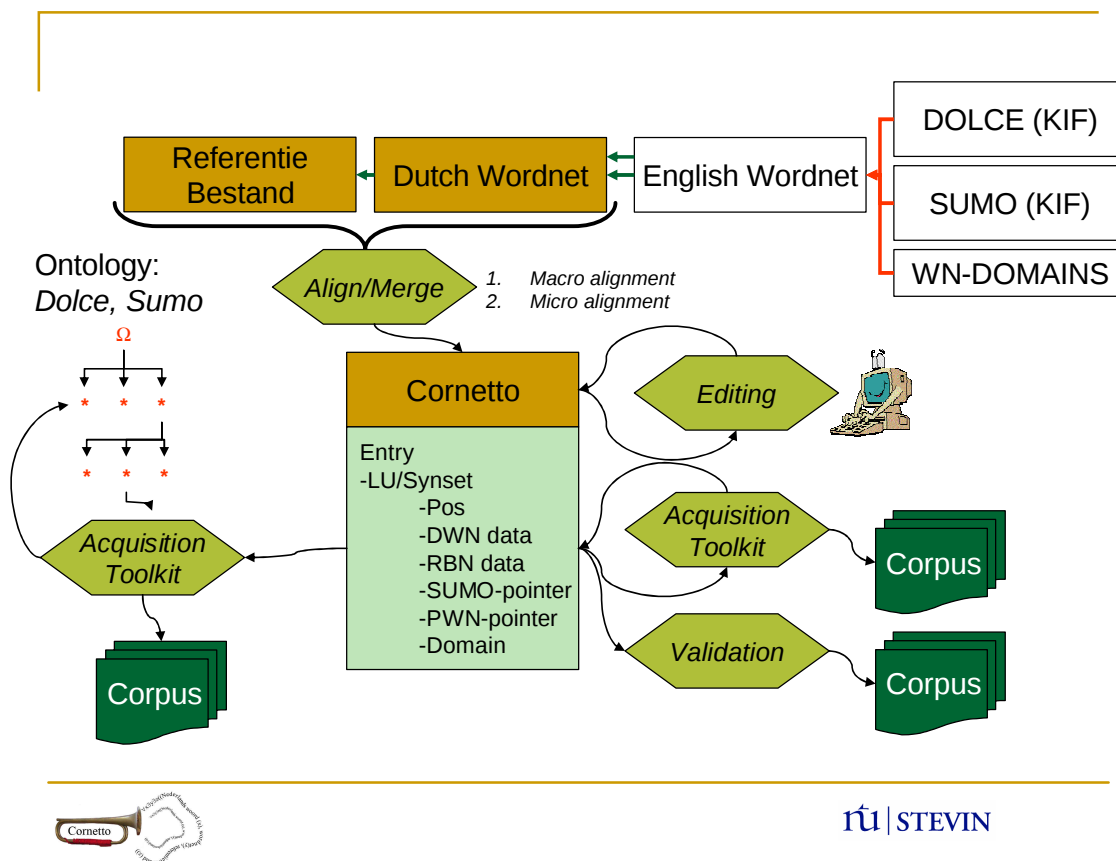


Figure 1 Overview of the Cornetto Project

Figure 1 gives an overview of the project. In the centre of this figure the Cornetto database that was created by merging two already existing Dutch lexical databases RBN and DWN. Via the original Dutch Wordnet there is a link to English Wordnet by means of an equivalence relation for the majority of all the Cornetto entries. The ontology labels (SUMO) and the Wordnet domains are imported through these equivalence relations to the English Wordnet. Possibly, DOLCE can be imported through these same relations but this has not been done in the project. The details on merging and aligning RBN and DWN into the initial Cornetto database will be further discussed in chapter 4. See also deliverable D02 for more details.

A considerable amount of time in this project was spent on editing the Cornetto database. Also, two acquisition toolkits were developed: one for acquiring new concepts and relations from corpora, the other for tuning and customizing to the legal-financial domain. More



information about the possibilities of these toolkits can be found in deliverables D06, D07, D08, D09, D10, and D11.

We decided to deliver more data than originally promised (over 92K entries instead of 40K entries). The main motivation for this is that coverage of a lexical resource is an important aspect for NLP applications. Furthermore, we expect to further improve the data in future releases and now provide the basic administrative units for making these improvements and enrichments. In many cases, the additional data are very specific, occurring either only in RBN or in DWN. In that case, alignment is not possible but the basic information provided in either resource is still valuable. The price for delivering more data than we promised is that we have not been able to manually verify and correct all the data. Furthermore, certain relations and other important data are not given for these words. They may lack a hyperonym relation, morpho-syntactic details for the Lexical Units, definitions or an equivalence relation to English WordNet.

The original DWN database was derived from the Van Dale Lexicographic Information System (VLIS). This is a much larger repository of which a portion was published in DWN. However, since the RBN entries may match VLIS entries that are not part of the DWN selection, we have extended the coverage of DWN with remaining VLIS entries. To give an impression of the data, the next table (Table 1) gives a global overview of the total number of form units (FUs) that are on both original database and their potential overlap. As can be seen, Cornetto can at most have 37,956 entries if we would restrict ourselves to words that occur in both resources, assuming the ideal case that all meanings can also be mapped one-to-one. By taking the union of the words, we get a much richer database (more than 96K entries), part of which is very rich, manually confirmed and corrected and aligned, part of which has partial and not-confirmed data.

Table 1: Form units (FUs) in the original source databases

Both in RBN and DWN	37,956
Total number FUs in Cornetto	96,659

This gives us a larger volume of lexical data to start with, with the drawback that a proportion of the data is not validated. However, since the most frequent and polysemous words are included in both RBN and DWN, we expect that the added bulk data are relatively straight-



forward and provide easily matching cases. To enable differentiating the data we used different lexical unit Ids and synset Ids for words and concepts that are derived from VLIS in addition to DWN: identifiers that start with the prefix “n-“ are add-ons from VLIS, identifiers that start with “d-“ originate from DWN.

We also added a so-called Alignment-Quality-Label (AQL) for matching records between the lexical units and the synsets (see below for more details). The users of the database can use these labels to get an indication of the core-Cornetto database that has both rich lexical unit and synset data, and the extended data that is partially represented. For different purposes, different selection of the data can be made.

The core Cornetto database is then the part of the database that is manually revised and checked. The work done for this part involved:

1. Manually aligning the lexical units from RBN and DWN:
 - a. linking lexical units
 - b. splitting lexical units
 - c. merging lexical units
 - d. deleting lexical units
 - e. adding lexical units as synonyms to existing synsets
2. Adding essential information to the lexical units involved:
 - a. morpho-syntax
 - b. semantics, e.g. resumes and case frames
 - c. examples
3. Adding essential relations to the synsets:
 - a. hyperonym relations
 - b. other significant relations
4. Manually verifying or creating the mapping to English Wordnet2.0
5. Manually verifying or creating the mapping to the SUMO ontology
6. Manually verifying or creating the mapping to WordNet Domains

In total about 10,108 lexical units have been edited manually, corresponding to about 4,500 words. We list the full set of words in the appendix. Another set of 509 nouns with 8 or more



equivalences to Wordnet2.0 and 618 verbs with 5 or more equivalence relations have been manually revised in terms of step 4), 5) and 6). In many cases also the other meanings of these words have been checked. If synsets got too many mappings through the automatic mapping software, the relations are usually of low-quality and therefore also the import of the SUMO and WordNet Domain labels is unreliable. By this manually revision, all doubtful cases have been resolved. The full list of lexical units is also given in the appendix.



3 Statistics

As explained above the overall database consists of 3 repositories. The overall statistics for these repositories is given in Table 2 and is distributed over the different POS. For newly created synsets and lexical units, no part-of-speech information can be derived from the identifiers (it needs to be derived from the lexical-unit), therefore, we listed these in a separate column.

Table 2: Overview data Cornetto repositories

	ALL	NOUNS	VERBS	ADJECTIVES	ADVERBS	OTHERS
Synsets	70,370	52,845	9,017	7,689	220	599
Lexical Units	119,108	85,449	17,314	15,712	475	158
Lemmas (form+pos)	92,686	70,315	9,051	12,288	1,032	
Synonyms in synsets	103,762	75,475	14,138	12,914	408	827
CID records	104,564	76,545	14,214	13,132	483	190
Synonym per synset	1.47	1.43	1.57	1.68	1.85	1.38
Senses per lemma	1.12	1.07	1.56	1.05	0.40	n.a.

Compared to the original database, we have far more data in the Cornetto database. For example, the database has about 1.6 more synsets (70,370) and 1.4 as many word meanings (103,762) as in the original Dutch wordnet: 44,015 synsets and 70,201 word meanings. Distributions over the different parts-of-speech do not show any deviant numbers, which also holds for the polysemy ratio (1.29, fc. 1.25 in the original DWN) and the number of synonyms per synset (1.47, fc. 1.59 in the original DWN). This is similar to ratios reported for wordnets in other languages.



3.1 Statistics for Synsets

The next table gives the detailed statistics for the synset data. The main fields are the synonyms, the internal semantic relations (from Dutch synset to Dutch synset), the equivalence relations to English Wordnet2.0, definitions, Wordnet Domain mappings, Sumo mappings and finally the Base Concept labeling.

Table 3: Overall data fields for synsets

Synsets	70370
Synonyms	103762
InternalRelations	144314
EquivalenceRelations	86040
Definitions	35770
WordNet Domains mappings	93470
Sumo mappings	70097
Base Level Concepts	9456

Almost half of the synsets have definitions, which are derived from the resume fields of all the lexical units that are synonyms of a synset. All the relations are one-to-many relations thus exceeding the number of synsets.

Table 4 gives the distribution of the language internal synset relations, most of which are symmetric pairs (see below for more details).

HAS_HYPERONYM	67533	INVOLVED	116
HAS_HYPONYM	54384	INVOLVED_AGENT	649
HAS_XPOS_HYPERONYM	92	INVOLVED_DIRECTION	8
HAS_XPOS_HYPONYM	90	INVOLVED_INSTRUMENT	1551
XPOS_NEAR_SYNONYM	2053	INVOLVED_LOCATION	177
HAS_HOLONYM	278	INVOLVED_PATIENT	974
HAS_HOLO_LOCATION	276	INVOLVED_RESULT	305
HAS_HOLO_MADEOF	145	INVOLVED_SOURCE_DIRECTION	13
HAS_HOLO_MEMBER	232	INVOLVED_TARGET_DIRECTION	30
HAS_HOLO_PART	1252	CO_ROLE	25
HAS_HOLO_PORTION	83	CO_AGENT_INSTRUMENT	53
HAS_MERONYM	277	CO_AGENT_PATIENT	40
HAS_MERO_LOCATION	279	CO_AGENT_RESULT	46
HAS_MERO_MADEOF	148	CO_INSTRUMENT_AGENT	52
HAS_MERO_MEMBER	231	CO_INSTRUMENT_PATIENT	282



HAS_MERO_PART	1246	CO_INSTRUMENT_RESULT	88
HAS_MERO_PORTION	85	CO_PATIENT_AGENT	39
HAS_SUBEVENT	415	CO_PATIENT_INSTRUMENT	283
IS_SUBEVENT_OF	409	CO_RESULT_AGENT	46
ROLE	116	CO_RESULT_INSTRUMENT	86
ROLE_AGENT	647	STATE_OF	444
ROLE_DIRECTION	9	BE_IN_STATE	433
ROLE_INSTRUMENT	1544	CAUSES	1205
ROLE_LOCATION	177	IS_CAUSED_BY	1190
ROLE_PATIENT	967	MANNER_OF	18
ROLE_RESULT	305	IN_MANNER	19
ROLE_SOURCE_DIRECTION	13	XPOS_NEAR_ANTONYM	16
ROLE_TARGET_DIRECTION	30	NEAR_ANTONYM	1649
		NEAR_SYNONYM	1137
		FUZZYNYM	6
		XPOS_FUZZYNYM	18
		TOTAL	133316

Table 5 and Table 6 give similar distribution data for the equivalence relations and the domains.

Table 4: Distribution of synset relations

HAS_HYPERONYM	67533	INVOLVED	116
HAS_HYPONYM	54384	INVOLVED_AGENT	649
HAS_XPOS_HYPERONYM	92	INVOLVED_DIRECTION	8
HAS_XPOS_HYPONYM	90	INVOLVED_INSTRUMENT	1551
XPOS_NEAR_SYNONYM	2053	INVOLVED_LOCATION	177
HAS_HOLONYM	278	INVOLVED_PATIENT	974
HAS_HOLO_LOCATION	276	INVOLVED_RESULT	305
HAS_HOLO_MADEOF	145	INVOLVED_SOURCE_DIRECTION	13
HAS_HOLO_MEMBER	232	INVOLVED_TARGET_DIRECTION	30
HAS_HOLO_PART	1252	CO_ROLE	25
HAS_HOLO_PORTION	83	CO_AGENT_INSTRUMENT	53
HAS_MERONYM	277	CO_AGENT_PATIENT	40
HAS_MERO_LOCATION	279	CO_AGENT_RESULT	46
HAS_MERO_MADEOF	148	CO_INSTRUMENT_AGENT	52
HAS_MERO_MEMBER	231	CO_INSTRUMENT_PATIENT	282
HAS_MERO_PART	1246	CO_INSTRUMENT_RESULT	88
HAS_MERO_PORTION	85	CO_PATIENT_AGENT	39
HAS_SUBEVENT	415	CO_PATIENT_INSTRUMENT	283
IS_SUBEVENT_OF	409	CO_RESULT_AGENT	46
ROLE	116	CO_RESULT_INSTRUMENT	86
ROLE_AGENT	647	STATE_OF	444
ROLE_DIRECTION	9	BE_IN_STATE	433
ROLE_INSTRUMENT	1544	CAUSES	1205
ROLE_LOCATION	177	IS_CAUSED_BY	1190
ROLE_PATIENT	967	MANNER_OF	18



ROLE_RESULT	305	IN_MANNER	19
ROLE_SOURCE_DIRECTION	13	XPOS_NEAR_ANTONYM	16
ROLE_TARGET_DIRECTION	30	NEAR_ANTONYM	1649
		NEAR_SYNONYM	1137
		FUZZYNYM	6
		XPOS_FUZZYNYM	18
		TOTAL	133316

Table 5: Distribution of equivalence relations

EQ_SYNONYM	3148
EQ_NEAR_SYNONYM	80084
EQ_HAS_HYPERONYM	1270
EQ_HAS_HYPERNYM	725
EQ_HAS_HYPONYM	257
EQ_HAS_HOLONYM	130
EQ_HAS_MERONYM	65
EQ_ROLE	55
EQ_INVOLVED	47
EQ_CO_ROLE	8
EQ_IS_CAUSED_BY	32
EQ_CAUSES	23
EQ_BE_IN_STATE	34
EQ_IS_STATE_OF	7
EQ_HAS_SUBEVENT	6
EQ_IS_SUBEVENT_OF	6
EQ_UNSPECIFIED	143
TOTAL	86040

**Table 6: Distribution of Wordnet Domain tags**

acoustics	146	doctrines	65	meteorology	375	skiing	69
administration	1914	drawing	145	metrology	1783	soccer	22
aeronautic	246	earth	42	military	1530	social_science	1
agriculture	634	ecology	102	money	997	sociology	856
alimentation	55	economy	2521	mountaineering	26	sport	1251
anatomy	2131	electricity	643	music	1613	state	4
anthropology	1079	electronics	6	mythology	170	statistics	25
applied_science	147	electrotechnics	32	number	488	sub	1
archaeology	51	engineering	49	numismatics	37	surgery	17
archery	13	enterprise	746	occultism	74	swimming	69
architecture	276	entomology	191	oceanography	6	table_tennis	46
art	768	ethnology	26	optics	196	tax	181
artisanship	342	exchange	569	painting	158	telecommunication	615
astrology	35	factotum	16433	paleontology	2	telegraphy	11
astronautics	50	fashion	1642	pedagogy	789	telephony	55
astronomy	269	fencing	20	person	1573	tennis	75
athletics	78	fishing	187	pharmacy	471	theatre	431
atomic_physic	112	folklore	30	philately	4	theology	60
auto	300	football	76	philology	105	time_period	615
badminton	4	free_time	413	philosophy	303	topography	2
banking	142	furniture	1108	photography	229	tourism	1010
baseball	100	gas	3	physics	1264	town_planning	1106
basketball	28	gastronomy	2956	physiology	1345	transport	3149
betting	28	genetics	3	plastic_arts	11	tv	63
biology	711	geography	1444	play	684	university	710
body_care	227	geology	870	politics	1583	volleyball	2
book_keeping	129	geometry	298	post	173	wrestling	22
botany	2164	golf	95	psychoanalysis	38	zoology	2615
bowling	19	grammar	93	psychology	2906	zootechnics	268
boxing	77	heraldry	161	publishing	920		
building_industry	3985	history	636	pure_science	228		
card	101	hockey	14	quality	3244		
chemistry	2038	hunting	105	racing	102		
chess	34	hydraulics	112	radio	93		
cinema	171	industry	1951	radiology	9		
color	201	insurance	196	railway	22		
commerce	1382	jewellery	117	religion	1431		
computer_science	403	law	2009	rowing	3		
cricket	11	linguistics	1216	rugby	5		
cycling	71	literature	791	school	817		
dance	93	mathematics	645	sculpture	52		
dentistry	29	mechanics	1036	sexuality	314		
diplomacy	20	medicine	2539	showjumping	2		
diving	36	merchant_navy	1077	skating	60		



The final Table 7 gives the miscellaneous data for the synsets. The synset database is a complex structure since it has many connections within the data collection and outside the collection. Making such a large database consistent is not trivial. Furthermore, we included all available data, which also includes non-validated relations.

Table 7: Miscellaneous data for synsets

MISCELLANEOUS	
MISSING TARGETS	
NOUNS	1
VERBS	9
ADJECTIVES	11
ADVERBS	1
TOTAL	22
NO INTERNAL RELATIONS	
NOUNS	532
VERBS	3
ADJECTIVES	3517
ADVERBS	109
NEW SYNSETS	12
TOTAL	4173

There are 22 targets of relations are missing. That means usually that the target synset was removed but the relation has not been updated, also due to bugs in the early version of the editor. The last set consists of synsets without any relations to any other synsets. There are 4,173 synsets in total, most of which are adjectives and nouns imported from the VLIS database that came without relations.

These minor errors will be removed in a future update of the database. Due to the complexity of the database, such errors may always occur and users of the database are advised to take such structures into account or ignore them.



3.2 Statistics for Lexical Units

The next tables give the statistics for the lexical units. The tables are divided over the major data fields: morphology, syntax, semantics, pragmatics and examples. We only listed the data elements that have values.

Table 8: Example fields

example	87315
examples	99228
category	80890
form_example	87315
canonicalform	86322
text_category	36028
textualform	16008
semantics_example	42932
semantics_noun	84765
semantics_verb	17285
sem_gc_compl	4975
sem_gc_gramword	4612
sem_lc_collocator	18830
sem_meaningdescription	16914
sem_subtype_argument	4524
synset_list	2
sy_combi	75944
sy_combicat	99366
sy_combipair	99516
sy_combiword	99510
sy_subtype	41034
sy_type	86006
syntax_example	86071

Table 9: Pragmatics fields

pragmatics	8573
prag_chronology	446
prag_connotation	1085
prag_domain	0
prag_frequency	734
prag_geography	909
prag_origin	2507
prag_socGroup	276
prag_style	2263
prag_subj_gen	9

**Table 10: Data elements for syntax, semantics and morphology**

SYNTAX		SEMANTICS		MORPHOLOGY	
syntax_noun	83308	arg	9538	flex_conjugation	10537
syntax_verb	17255	args	5265	flex_conjugationtype	10537
syntax_adj	15633	caseframe	5393	flex_mode	10586
sy_advusage	8158	caserole	9534	flex_number	10590
sy_article	39147	selrestriction	8629	flex_pastpart	3120
sy_class	10626	selrestrole	9483	flex_pasttense	3121
sy_comp	17884	sem_caseframe	5476	flex_person	10593
sy_compl_text	2245	sem_countability	40421	flex_tense	10593
sy_complementation	13422	sem_def	18644	morphology_adj	15621
sy_gender	39086	sem_defSource	57401	morphology_noun	83357
sy_number	504	sem_def_noun	39076	morphology_verb	17235
sy_peraux	10613	sem_definition	60476	mor_base	1268
sy_position	8161	sem_genus	39076	mor_comparative	4274
sy_reflexiv	10613	sem_reference	39133	mor_comparis	8149
sy_separ	10572	sem_resume	59467	mor_comparison	8148
sy_subject	10586	sem_selrestriction	8632	mor_declinability	8147
sy_trans	10627	sem_selrestrictions	3946	mor_flectional_type	1624
sy_valency	10627	sem_shift	10793	mor_superlative	4275
		sem_spec_collocator	7112	morpho_plurform	30651
		sem_specifcae	36618	morpho_plurforms	29043
		sem_subclass	5182	morpho_structure	13309
		sem_type	57951	morpho_type	57939
		semantics_adj	15703		

3.3 Statistics for CID records

In Table 11, we give the CID records that match lexical units to synsets and have been selected. There are many more CID records that give lower rankings of matches and that are not selected.

Table 11: Number of matching and non-matching Lexical Units

DWN and RBN matches	35,289	37.74%
LUs only in DWN	54,983	58.81%
LUs only in RBN	3,223	3.45%
Total	93,495	

About 38% of all these records are matches between a lexical unit in DWN and a lexical unit in RBN. Almost 60% consists of lexical units from DWN that are not matched with RBN. These are mostly words not occurring in RBN. For each of these, we created a so-called



dummy lexical unit in the LU repository with minimal information. Similarly, 3,223 lexical units from RBN could not be matched with a lexical unit in DWN. For these, we did not create a new synset. The reason for this is that they often can be added as a synonym to an existing synset.

The final two tables give information on the status of the mappings. In Table 12, we see that about 53% of the CID records have no value for status. This means that none of the editors has looked at the matches and they have not been validated in a post selection. Most of these are nouns that only occur in DWN as a word.

The records with a status field are differentiated for their so-called AQL. These labels are explained in more detail below. Briefly, the non-manual mappings are post-processed using different rules, where the number suffix indicates the reliability of the process, based on a sample of 100 records per part of speech. For example, M-97 is a mapping for a monosemous word in RBN to a monosemous words in DWN with a confidence of 97%. AQL values are: B-95, BM-90, D-55, D58, D-75, M-97 and RESUME-75.

Table 12: Status values for CID records

No status value	55,975	53.53%
Status value	48,589	46.47%
manual	10,120	9.68%
B-95	4,944	4.73%
BM-90	4,214	4.03%
D-55	171	0.16%
D-58	774	0.74%
D-75	2,085	1.99%
M-97	25,234	24.13%
RESUME-75	1,047	1.00%
TOTAL	104,564	

Another field is used to store the author for the mapping. Table 13 indicates that more than 8% of all the mappings have an author label.

Table 13: Author labels for CID mappings

authored	9,526	9.11%
automatic	95,038	90.89%
TOTAL	104,564	



4 Database Design and Alignment

4.1 Design

The Cornetto database (CDB) consists of 3 main data collections:

- Collection of Lexical Units, mainly derived from the Referentiebestand Nederlands (RBN)
- Collection of Synsets, mainly derived from Dutch Wordnet (DWN)
- Collection of Terms and axioms, mainly derived from SUMO and MILO

Both DWN and RBN are semantically based lexical resources. RBN uses a traditional structure of form-meaning pairs, so-called Lexical Units (Cruse1986). Lexical Units are word senses in the lexical semantic tradition. They contain all the necessary linguistic knowledge that is needed to properly use the word in a language. Word meanings that are synonyms are separate structures (records) in RBN. They have their own specification of information, including morpho-syntax and semantics. For more information about RBN, please refer to RBN-Documentatie (Martin e.a., 2005).

DWN is organized around the notion of Synsets. Synsets are concepts as defined by Miller and Fellbaum (Miller e.a. 1991, Fellbaum 1998) in a relational model of meaning. They are mainly conceptual units strictly related to the lexicalization pattern of a language. Concepts are defined by lexical semantic relations.¹ Typically in Wordnet, information is provided for the synset as a whole and not for the individual word meanings. For example, in Wordnet the synset has a single gloss but the different lexical units in RBN each have their own definition. From a Wordnet point of view, the definitions of lexical units that belong to the same synset should thus semantically be compatible or synonymous. For more information about DWN, please refer to EuroWordnet General Document, Vossen (2002).

Outside the lexicon, an ontology will provide a third layer of meaning. The Terms in an ontology represent the distinct types in a formal representation of knowledge. Terms can be combined in a knowledge representation language to form expressions of axioms. In principle, meaning is defined in the ontology independently of language but according to the principles of logic. In Cornetto, the ontology represents an independent anchoring of the

¹ For Cornetto, the semantic relations from EuroWordNet are taken as a starting point (Vossen1998).



relational meaning in Wordnet. The ontology is a formal framework that can be used to constrain and validate the implicit semantic statements of the lexical semantic structures, both the lexical units and the synsets. In addition, the ontology provides a mapping of a vocabulary to a formal representation that can be used to develop semantic web applications. For more information about SUMO please refer to <http://www.ontologyportal.org/>.

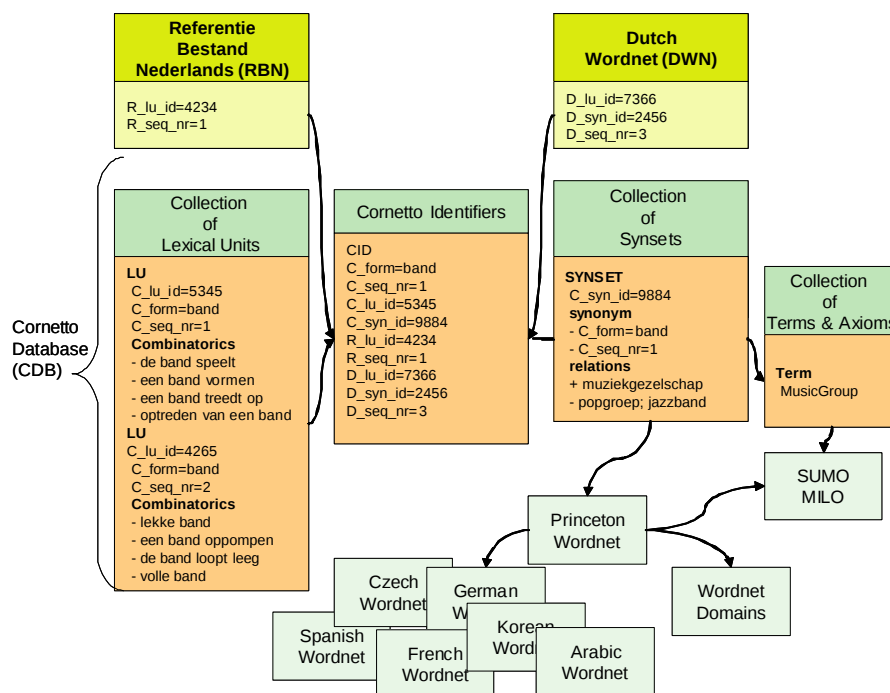


Figure 2 Data collections in the Cornetto Database

Figure 2 shows an overview of the different data structures and their relations. The different data can be divided into 3 layers of resources, from top to bottom:

- The RBN and DWN (at the top): the original database from which the data are derived;
- The Cornetto database (CDB): the ultimate database;
- External resources: any other resource to which the CDB will be linked, such as the Princeton Wordnet, wordnets through the Global Wordnet Association, Wordnet domains, ontologies, corpora, etc.

The center of the CDB is formed by the table of CIDs. The CIDs tie together the separate collections of LUs and Synsets but also represent the pointers to the word meaning and



synsets in the original databases: RBN and DWN and their mapping relation. The CIDs are just administrative records of the so-called Cornetto Identifiers (CIDs). These identifiers contain the relations between the lexical units and the synsets in the CDB but also to the original word senses and synsets in the RBN and DWN (and VLIS if applicable).

A CID record contains the following information:

cid_id	= the identifier of the cid record
form	= form of the word in Cornetto
pos	= Cornetto part-of-speech
seq_nr	= the sequence of sense number in Cornetto
c_lu_id	= the identifier of the lexical unit in Cornetto
c_sy_id	= the identifier of the synset in Cornetto
cid score	= alignment score of the first automatic alignment procedures (see 4.2)
name	= name of the editor in case of manual alignment
status	= alignment quality label (see 4.2)
selected	= <i>not relevant anymore</i>
d_syn_id	= the identifier of the of the synset in DWN from which it was derived
d_lu_id	= the original identifier of the lexical unit in DWN
d_seq_nr	= the original sense sequence number or sense number in DWN
r_lu_id	= the original identifier of the lexical unit in RBN
r_seq_nr	= the original sense sequence number in RBN
r_lu_id	= the identifier of the lexical unit in RBN from which it was derived
r_seq_nr	= the original sequence number or sense number in RBN



4.2 Alignment Quality Label

The alignment of the RBN Lexical Units and the DWN synsets has been performed in three steps:

- Automatic alignment

Automatic Alignment of all words has been done following a number of mapping strategies which use the overlapping features between the two resources². The result is a mapping table (cid) representing the alignment, where each word meaning in RBN is mapped to one of more word meanings of the same word (with the same part-of-speech) in DWN. The mappings are scored; these scores can be found in the cid table (cid_score). Note that the mapping procedure generated new sense number for the lexical units, which are Cornetto sense numbers. Cornetto sense numbers are in principle the same as the RBN sense numbers but additional sense numbers can be created if DWN senses are not aligned. the extra DWN senses then have sequential numbers starting from the highest RBN sense.

- Manual Alignment

The most frequent and most polysemous words have been aligned manually; the results of the automatic alignment formed the input for the editing work.

- Automatic (re-)alignment

Finally some extra alignments strategies have been carried out with regard to:

- (1) monosemous and bisemous words and
- (2) words with a considerable overlap between the RBN and DWN definitions

The different alignment mappings are assessed by determining accuracy at a sample of 100 alignments per type. This resulted in an alignment quality label (AQL) which is stored in the in the field *status* of the cid records.

² For more information see CornettoDeliverableD02
(<http://www.let.vu.nl/onderzoek/projectsites/cornetto/deliver.html>)



Table 14: Alignment quality labels

AQL	Accuracy	Pos	Number of alignments
Manual³	100%		10120
		A	1178
		V	1836
		N	7099
		Adv	7
M-97	97%		25234
		A	4056
		V	2873
		N	18305
B-95	95%		4944
		A	596
		V	1663
		N	2685
BM-90	90%		4214
		A	733
		V	896
		N	2584
		Adv	1
Resume-75	75%		1047
		A	121
		V	846
		N	80
D-75	75%	N	2085
D-58	97%	V	774
D-55	95%	A	171
NO-STATUS	n.a.		55975
		A	6320
		V	5326
		N	43741
		Adverb	476
		Pronoun	13
		Interjection	80
		Conjunction	6
		Numeral	6
		Preposition	7

³ Appendix B gives the complete list of manually aligned words (about 4,500 in total)



- **Manual** includes:
- frequent verbs (CGN -freq > 250) ⁴
- frequent adjectives (CGN-freq > 300)
- highly polysemous nouns (number of senses -> 3 in Cornetto after the alignment)
- highly polysemous adjectives (number of senses > 3 in RBN or DWN)
- highly polysemous verbs

M-97 includes:

- monosemous words, i.e. words that have in both DWN and RBN one sense only

B-95 includes:

- bisemous words, i.e. words that have both in DWN and RBN two senses. The first sense and the second sense of RBN are aligned with the first sense and the second sense of DWN, respectively,

BM-90 includes:

- words that have one sense in RBN and two senses in DWN or viceversa. The first sense of the bisemous word is aligned with the only sense of the monosemous word. The second sense remains unaligned (which is usually correct).

Resume-75 includes:

- Lexical Unit – Synset alignments based upon a substantial overlap between the RBN and DWN definitions and synonyms.

D-75 includes:

- Noun Lexical Unit – Synset alignments that have an automatic alignment score >30%.

D-58 includes:

- Verb Lexical Unit – Synset alignments that have an automatic alignment score >30%.

D-55 includes:

- Adjective Lexical Unit – Synset alignments that have an automatic alignment score >30%.

No-status

- Lexical Unit- Synset alignments between a DWN synset and a dummy Lexical Unit.

Dummy lexical units are created for those DWN word meanings that could not be mapped to RBN lexical units automatically. As can be seen from Table 11 and Table 14, these represent

⁴ CGN = Corpus Gesproken Nederlands



a large proportion of the lexical units without status information: 48,580 out of 56,122 lexical units of which most are nouns. These LUs have only minimal morpho-syntactic information and no semantics, pragmatics and examples. As explained above, the suffix of the AQL indicates the confidence of the mapping based on samples of 100 records differentiated per part-of-speech. The AQL D-75 thus means it is correct with a confidence of 75%.



5 Data Collections

In this chapter, we give some short descriptions of two main data collections in the Cornetto database: the synsets (5.1) and the lexical units (5.2). For extensive information on lexical units and synsets, see the RBN Documentation (Martin e.a., 2005) and the Wordnet documentation (EuroWordnet (Vossen, 2002)).

5.1 Cornetto Synsets

The Cornetto synset database is organised around synsets and their semantic relations. A synset is a set of synonyms, i.e. words with the same part of speech that can be interchanged in a certain context. For example {car, motorcar, automobile, machine} form a synset because each word can be used to refer to the same concept. The synsets are related to each other by semantic relations like hyperonymy, hyponymy. Note that some semantic relations are used across part of speech.

The meaning of the synset is established by the set of synonyms and the relations to other synsets. Nevertheless, synsets can also have definitions (about 50% of the database). These definitions are amalgamated from the resume-fields of the lexical units that make up the synonyms. Other ways of defining the meaning are provided by the domain label and the mapping to Sumo. Finally, some synsets are marked as Base Concepts: the set of concepts that are used to build many different wordnets in the world (see the general EuroWordNet documentation for further details, Vossen 2002).

In the next sections we will discuss the main characteristics and structures for nouns, verbs and adjectives.

5.1.1 Nouns

5.1.1.1 ID number

Each synset has a unique Cornetto identification number (c_sy_id in xml). The Cornetto ID number of a synset is the same as the ID of the synsets in the original Dutch Wordnet, unless the synset was created manually in Cornetto.



Each ID number can start with a ‘d_’ (the synset originates from the original Dutch Wordnet), an ‘n_’ (the synset originates from Vlis (see above) or a ‘c_’ (the synset is new and created in the Cornetto project). In the case of nouns, these codes are followed by an ‘n-’ denoting the part of speech of the synset. These two letters are followed by the unique number. For example:

d_n-34987 (attribute in xml: c_sy_id="d_n-34987", d_synset_id="d_n-34987")

n_n-504746 (attribute in xml: c_sy_id="n_n-504746", d_synset_id="n_n-504746")

c_96 (attribute in xml: c_sy_id="c_96")

Note that synsets starting with a ‘c_’ are directly followed by the ID number, i.e. the part-of-speech tag is not represented in the ID.

5.1.1.2 Synonyms

Each synset contains at least one lexical unit, but often more⁵. The lexical units *in* the synset have been aligned with the lexical units in the RBN database. This has some consequences on the information found in the list of synonyms in the synset. There are five types of synonym description in the XML that denote the status of the alignment. The next example comes from the large synset *ontlasting* (d_n-37085) meaning ‘excrement/faeces’. In bold three of the lexical units are listed from this synset, all with a different status: *uitwerpsel*:1, *drek*:1 and *beer*:3.

{*excrement*}

- A) <synonym status="" c_cid_id="130975" c_lu_id-previewtext="**uitwerpsel:1**" c_lu_id="d_n-304410"/>
- B) <synonym status="" c_cid_id="12386" c_lu_id-previewtext="**drek:1**" c_lu_id="r_n-11508"/>
- C) <synonym c_lu_id-previewtext="**beer:3**" c_lu_id="r_n-6271"/>
- D) <synonym c_lu_id-previewtext="simulated:x" c_lu_id="c_543210"/>
- E) <synonym c_lu_id-previewtext="simulated:y" c_lu_id="d_523456"/>

Synonym A starts with an empty synonym status attribute, followed by a c_cid_id that points to the ID number of the automatic alignment between the word *uitwerpsel* in the synset and the word *uitwerpsel* in the lexical unit. The c_lu_id at the end is followed by an identification

⁵ Except for the miscellaneous cases discussed in section 3.1



number that starts with a 'd_'. This means that the word *uitwerpsel* from the synset 'excrement/faeces' could not be aligned with an existing lexical unit in the RBN database, and therefore, an LU id from DWN was used. This results in a dummy lexical unit in the Cornetto lexical units. Note that the editors had the choice to complete this initial empty information with further information (E) or to create a new lexical unit (D).

Synonym B has the same structure as described above, but the `c_lu_id` is followed by an ID number starting with a 'r_' which means that this word was aligned with an existing RBN lexical unit. However, in all of these cases that doesn't imply that the automatic alignment with this lexical unit will be correct. The void status attribute indicates that these are not manually checked or created.

Synonym C differs from the other two: there is no synonym status or `c_cid_id`. In this case, the automatic alignment was checked by hand and approved. Note that the `c_lu_id` in this example can also be followed by ID numbers starting with a 'c_' (see:D) or with a 'd_' (see:E). In the first case the editor reused a dummy lexical unit, in the last case the editor created a new Cornetto lexical unit.

Synonym D is simulated here to complete this overview. In this case a new lexical unit was created in the RBN lexical units and matched to this synset. This type of lexical unit starts with 'c_' and is followed by a six digit number that always begins with a '54'. Note that there is no `c_cid_id` or synonym status.⁶

Synonym E is also simulated; it shows a mapping with an automatically derived dummy lexical unit that was further elaborated by an editor and manually confirmed. Note again that there is no that there is no `c_cid_id` or synonym status.

The next examples show synonyms that have status fields that correspond to the ACQL labels in the CID records (discussed above). The status `rbn-1-dwn-1` means that a momosemous RBN word was matched to a monosemous DWB word. In the case of `rbn-2-dwn-2`, the word had two meanings, and `rbn-dwn-1-2` means that it had one meaning in one resource and two in

⁶ The status fields should have been filled automatically by the editor but this has not been implemented.



the other. The status RESUME MATCH indicates that it is derived from overlap in definition words and synonyms.

```
<synonym c_cid_id="17678" c_lu_id="r_n-15326" c_lu_id-previewtext="gluurder:1" status="rbn-1-dwn-1"/>
<synonym c_cid_id="54337" c_lu_id="r_n-41758" c_lu_id-previewtext="voyeur:1" status="rbn-1-dwn-1"/>
<synonym c_cid_id="2933" c_lu_id="r_n-3091" c_lu_id-previewtext="afzetting:3" status="RESUME
MATCH"/>
<synonym c_cid_id="5795" c_lu_id="r_n-6786" c_lu_id-previewtext="beroepsgeheim:1" status="rbn-1-dwn-
1"/>
<synonym c_cid_id="" c_lu_id="r_n-3414" c_lu_id-previewtext="ambtsgeheim:1" status="rbn-2-dwn-2"/>
<synonym c_cid_id="18956" c_lu_id="r_n-16336" c_lu_id-previewtext="hardrijder:1" status="rbn-dwn-1-2"/>
<synonym c_cid_id="" c_lu_id="c_545911" c_lu_id-previewtext="lijden:3" status="manual"/>
<synonym c_cid_id="" c_lu_id="c_545506" c_lu_id-previewtext="wacht:2" status="manual"/>
```

The precise status is always records in the CID record rather than the synset structure. It is stored here only for ease of reference. For example, the status manual is only occasionally stored here but consistently in the CID table.

5.1.1.3 Internal Relations

All synsets are connected to each other by semantic relations. Every noun synset has at least one basic semantic relation: the hyperonym relation. An overview of the relations is given in Table 4 in section 3. For more information about the semantic relations, please refer to EuroWordnet General Document, Vossen (2002).

Semantic relations in Cornetto can have three slightly different representations in the XML. The first synset of Example (1) (auto: 1) shows an already existing semantic relation in the database. The second synset (beer:5) shows a new semantic relation - created within Cornetto; the third shows an automatically implied relation that points back to the initial relation made in beer:5. If an internal relation is created from synset A to synset B; there is always an automatically generated relation from B back to A. From the synset 'buigen:3, krommen:4' it can be seen that these relations are marked with the tag 'relation-reversed = true'.



example (1)

auto:1/d_n-26965

<wn_internal_relations>

<relation factive="false" reversed="true" negative="false" relation_name="ROLE_PATIENT" target-previewtext="tectyleren:1" coordinative="false" disjunctive="false" target="d_v-6948">

<author score="0.0" status="YES" name="Piek" date="19961217" source_id="d_n-26965"/>

beer:5/c_94

<wn_internal_relations>

<relation factive="false" reversed="false" negative="false" relation_name="ROLE_INSTRUMENT" target-previewtext="buigen:3, krommen:4" coordinative="false" disjunctive="false" target="d_v-1699"/>

buigen:3, krommen:4/d_v-1699

<wn_internal_relations>

<relation reversed="true" target-previewtext="beer:5" relation_name="INVOLVED_INSTRUMENT" generated="true" target="c_94"/>

All internal semantic relations can be found within the element <wn_internal_relations>. The semantic relations have five characteristics that can have the value true or false: factive, reversed, negative, coordinative and disjunctive. The default value is false. A short illustration of this is to be found in the synset ‘auto:1’ (of example (1)), where we find a reversed ROLE_PATIENT relation with the verb ‘tectyleren’ (to undercoat a car to prevent it from corrosion). This means that the relation from ‘tectyleren’ to ‘auto’ is stronger than from ‘auto’ to ‘tectyleren’. More information on this subject can be found in the EuroWordnet documentation. The semantic relation tag is followed by a target preview text (tectyleren:1) and the according ID number of the target synset (d_v-6948). After this ID number, there is a line with author information that originates from the Dutch Wordnet database. The names in the author fields are for the human editors that created the relation, except for “Paul”, which stands for automatic relations or alignment. Automatically created links can still be confirmed by a human editor, in which case the status field is set to “true”. The source_id at the end of the author element, is the number of the actual synset (‘auto:1’). The newly created semantic relations lack information about the author score, status, name, date and source_id.⁷

⁷ Again, this functionality was not implemented in the editor.



If an internal relation is created from synset A to synset B; there is always an automatically generated relation from B back to A. This means that almost each internal semantic relation has its counterpart, e.g. hyperonym-hyponym. This is illustrated by the synsets ‘beer:5’ and ‘buigen:3, krommen:4’ (example 2). In the case of ‘beer:5’, there is a `ROLE_INSTRUMENT` relation to ‘buigen:3, krommen:4’. The synset ‘buigen:3, krommen:4’ has an `INVOLVED_INSTRUMENT` relation to beer:5. This relation has been created automatically by making the relation from ‘beer’ to ‘buigen:3, krommen:4’. Note that in these cases, there is an automatically generated ‘reversed’ label in the implied relation.

Furthermore, these implied relations have a default value ‘false’ for the features factive, negative, coordinative and disjunctive, but these defaults are not specified in the XML.

example (2)

```
<relation_name="ROLE_INSTRUMENT" target-previewtext="buigen:3, krommen:4, krommen:2"
coordinative="false" disjunctive="false" target="d_v-1699"/>

<relation reversed="true" target-previewtext="beer:5" relation_name="INVOLVED_INSTRUMENT"
generated="true" target="c_94"/>
```

5.1.1.4 Wordnet Equivalence relations

The Cornetto synsets have been mapped automatically to Princeton Wordnet 2.0 and 3.0 (PWN); the quality of these equivalents mappings is indicated by a score. Each equivalence relation starts with a relation name; in most cases this will be an `EQUAL_NEAR_SYNONYM`, an `EQUAL_SYNONYM` or an `EQUAL_HAS_HYPERONYM` relation. An overview of these relations is given in

HAS_HYPERONYM	67533	INVOLVED	116
HAS_HYPONYM	54384	INVOLVED_AGENT	649
HAS_XPOS_HYPERONYM	92	INVOLVED_DIRECTION	8
HAS_XPOS_HYPONYM	90	INVOLVED_INSTRUMENT	1551
XPOS_NEAR_SYNONYM	2053	INVOLVED_LOCATION	177
HAS_HOLONYM	278	INVOLVED_PATIENT	974
HAS_HOLO_LOCATION	276	INVOLVED_RESULT	305
HAS_HOLO_MADEOF	145	INVOLVED_SOURCE_DIRECTION	13
HAS_HOLO_MEMBER	232	INVOLVED_TARGET_DIRECTION	30
HAS_HOLO_PART	1252	CO_ROLE	25



HAS_HOLO_PORTION	83	CO_AGENT_INSTRUMENT	53
HAS_MERONYM	277	CO_AGENT_PATIENT	40
HAS_MERO_LOCATION	279	CO_AGENT_RESULT	46
HAS_MERO_MADEOF	148	CO_INSTRUMENT_AGENT	52
HAS_MERO_MEMBER	231	CO_INSTRUMENT_PATIENT	282
HAS_MERO_PART	1246	CO_INSTRUMENT_RESULT	88
HAS_MERO_PORTION	85	CO_PATIENT_AGENT	39
HAS_SUBEVENT	415	CO_PATIENT_INSTRUMENT	283
IS_SUBEVENT_OF	409	CO_RESULT_AGENT	46
ROLE	116	CO_RESULT_INSTRUMENT	86
ROLE_AGENT	647	STATE_OF	444
ROLE_DIRECTION	9	BE_IN_STATE	433
ROLE_INSTRUMENT	1544	CAUSES	1205
ROLE_LOCATION	177	IS_CAUSED_BY	1190
ROLE_PATIENT	967	MANNER_OF	18
ROLE_RESULT	305	IN_MANNER	19
ROLE_SOURCE_DIRECTION	13	XPOS_NEAR_ANTONYM	16
ROLE_TARGET_DIRECTION	30	NEAR_ANTONYM	1649
		NEAR_SYNONYM	1137
		FUZZYNYM	6
		XPOS_FUZZYNYM	18
		TOTAL	133316

Table 5 in section 3.

The equivalence relations have different indications and ranges of quality scores depending on the origin and date of the equivalence mappings. In the first synset of (3, an older mapping from Dutch Wordnet to Princeton Wordnet 1.5 (ENG15-00033941-n) was re-mapped to the 2.0 edition (ENG20-00059106-n). The latter mappings have again been extended with a mapping to WordNet3.0 (ENG30-00064504-n). The original score for this mapping is found after the author score and is in this case 670.0. The scores for these mappings vary from 500, as being the lowest score, up to 10.000, as the highest. These score have not been normalized.

In the second synset of example ((3), the original mapping to PWN1.6 was converted to PWN2.0, which was again converted to PWN3.0. The score in this type of equivalence relations is ranked from 42 as the lowest score up to 100 as the highest. These scores have been normalized.

The last synset of example ((3) gives an equivalence relation that is not derived automatically but that is manually created within the Cornetto project. The automatically derived relations



have been removed and a new equivalence relation to PWN was created. Note that ‘version’ is empty and that there is no score.

example (3)

```
(d_n-25553, topper:1)
<wn_equivalence_relations>
relation relation_name="EQ_NEAR_SYNONYM" target20-previewtext="hit:3, smash:5, smasher:3, strike:6,
bang:5" version="pwn_1_5" target30="ENG30-00064504-n" target20="ENG20-00059106-n" pos=""
target="ENG15-00033941-n">
<author score="670.0" status="" name="Paul" date="19970903"

(d_n-14596, topper:3)
<wn_equivalence_relations>
relation relation_name="EQ_NEAR_SYNONYM" target20-previewtext="colossus:2, behemoth:2, giant:2,
heavyweight:5, titan:1" version="pwn_1_6" target30="ENG30-09938991-n" target20="ENG20-09304064-n"
pos="" target="ENG16-07167525-n">
<author score="42.0" status="" name="Iרון Technologies" date="20070622" source_id=""/>

(d_n-14596, topper:3)
<wn_equivalence_relations>
relation relation_name="EQ_NEAR_SYNONYM" target20-previewtext="best:2, topper:3" version=""
target30="ENG30-09851165-n" target20="ENG20-09223355-n" pos="">
<author score="" source_id=""/>
```

Only the last synset of the above example was checked by hand. Manual editing has been done for most of the manually checked alignments as well. Additionally, we have manually edited all nominal synsets that have more than 8 equivalence relations and all verbal synsets that have more than 4 equivalence relations (see the appendix for a full list of these synsets).

5.1.1.5 Ontology

The SUMO ontology terms were imported through the automatically derived equivalence relations to PWN2.0 (Niles and Pease 2001, Niles and Pease 2003). We copied the mapping relations of the English synsets to SUMO and MILO to the equivalent synsets in Dutch. Mappings were further manually adapted in the editing process.

In the SUMO to PWN mapping, the following relations are used:



- = the synset is equivalent to the SUMO concept
- + the synset is subsumed by the SUMO concept
- @ the synset is an instance of the SUMO concept
- [the SUMO concept is subsumed by the synset

For an overview and description of all relations and terms, see the documentation and the knowledge browser on: <http://www.ontologyportal.org>.

The mappings from PWN to SUMO are binary relations. In Cornetto, we used triplets to make more complex expressions. For the triplets, the above relations have been extended with all the relations that were defined in SUMO (April 2006) and are used in the axioms. A full list of these relations is given in Table 15.

Table 15: SUMO relations used in Cornetto

age	agent	ancestor	attends
attribute	before	beforeOrEqual	believes
bottom	brother	causes	causesSubclass
citizen	completelyFills	component	conclusion
connected	considers	consistent	contains
containsInformation	cooccur	crosses	date
daughter	desires	destination	diameter
direction	disjoint	distributes	duration
during	earlier	element	employs
entails	equal	exactlyLocated	experiencer
exploits	faces	familyRelation	father
fills	finishes	geographicSubregion	geometricPart
geopoliticalSubdivision	grasps	hasPurpose	hasSkill
height	holdsDuring	hole	home
husband	inhabits	inhibits	instance
instrument	interiorPart	inverse	knows
larger	leader	legalRelation	length
lineMeasure	located	manner	material
measure	meetsSpatially	meetsTemporally	member
modalAttribute	monetaryValue	mother	needs
origin	overlapsPartially	overlapsSpatially	overlapsTemporally
parallel	parent	part	partiallyFills



partlyLocated	path	patient	penetrates
piece	possesses	prevents	properPart
properlyFills	property	realization	refers
represents	resource	result	sibling
side	sister	smaller	son
spouse	starts	stays	subclass
subset	surface	temporalPart	time
Top	transactionAmount	traverses	true
Uses	wants	wears	width
Wife			

The SUMO element ‘ont_relation’ contains a relation attribute (relation_name) with either the characters ‘+’, ‘=’, ‘[’, ‘@’ or one of the above relations as a value. Furthermore it has two attributes (arg1 and arg2) that represent the two arguments of the triplets. Other attributes are status ‘true’ or ‘false’, a name for whom created the mapping, and an attribute for ‘negative’. Below is an example of the ‘ont_relation’ element:

```
<ont_relation status="false" name="dwn10_pwn16_pwn20_mapping" negative="false" relation_name="+"
arg1="" arg2="Artifact"/>
```

Depending on the ontology mapping being checked or created by hand or not, the XML of the SUMO relations may vary.

The relation name and the two arguments represent the triplet. The triplets are used as simplified representations of semantic implications. The arguments of the triplets follow the syntax of the relation names in SUMO. The fillers can be a SUMO term or a variable. Variables are integers, where the integer ‘0’ is reserved to co-index with the referent of the synset that is being related. Empty argument slots are assumed to hold the value ‘0’ as well. For example the following expressions are possible in the Cornetto database, where each a) and b) example is equivalent:

1. Equality:
 - a. (=, 0, Circle)
 - b. (=, , Circle)
2. Subsumption:



- a. (+, 0, Artifact)
 - b. (+, , Artifact)
- 3. Related:
 - a. (part, 0, PlantBranch)
 - b. (part, , PlantBranch)
- 4. Axiomatized:
 - a. (instance, 0, Water) (instance, 1, Making) (instance, 2, Tea) (resource, 0, 1) (result, 2,1)
 - b. (instance, , Water) (instance, 1, Making) (instance, 2, Tea) (resource, , 1) (result, 2,1)

Most of the relations imported from the English Wordnet will have the structure of 1b and 2b. The other triplets are used instead of the complex SUMO-KIF expressions in SUMO. The triplets are used to specify a complex mapping relation to the SUMO ontology, in case the basic mapping relations are not sufficient. This is especially the case for so-called non-rigid concepts that are not present in SUMO, e.g. ‘waakhond’ (watchdog) is not a type of dog but a dog in a particular role (Guarino and Welty 2002). The following simplified expression can then be found in the Cornetto database for the non-rigid synset of {waakhond}:

(instance, 0, Canine) (instance, 1, Guarding) (role, 0, 1)

By assuming default values for the KIF syntax, it is possible to generate more complex expressions that come close to the axioms in SUMO. Such an interpretation can be derived as follows: the default operator for a list of the triplets is AND, and we assume default existential quantification of any of the variables, specified as a value of the arguments. Furthermore, we follow the convention to use a zero symbol as the variable that corresponds to the denotation of the synset being defined and any other integer for other denotations. Finally, we use the symbol $\Leftrightarrow$ in our explanation of the triplets for full equivalence (bidirectional subsumption). In the case of partial subsumption, we use the symbol $\Rightarrow$, meaning that the KIF expression is more general than the meaning of the synset. If no symbol is specified, we assume an exhaustive definition by the KIF expression. The symbol $\Leftrightarrow$ applies by default. The above triplet for ‘waakhond’ should then be read as follows:



The expression exhaustively defines the synset ($\hat{U}$), AND there exists an instance 0 of the type Canine (instance, 0, Canine), AND any referent of an expression with the synset {waakhond} as the head is also an instance of the type Canine (the special status of the zero variable), AND there exists an instance of the type Guarding 1 (instance, 1, Guarding), AND the entity 0 has a role relation with the entity 1 (role, 0, 1).

The triplets can also be used to define new rigid types that are not in SUMO. In that case, we state that the synset has the names for these types. For names of rigid types, e.g. ‘hond’ as a Dutch name for Canine, we propose the following expressions in Cornetto:

hond (=, 0, Canine); the synset {hond} is a Dutch name for the rigid type Canine
 bokser (+, 0, Canine); the synset {bokser} is a Dutch name for a rigid concept which is a subclass of the type Canine

Naming relations are mostly imported from the SUMO mappings to the English Wordnet through the equivalence relation of the Dutch synset to the English synset. In the case of {bokser}, the mapping needs to be manually added because this dog race is not in the English Wordnet and not in SUMO. Possibly, SUMO could be extended with this type.

Within the wordnet hierarchy, we find many cases of mixtures of rigid and non-rigid concepts that have the same hyponym relation, i.e. both ‘waakhond’ (watchdog) and ‘bokser’ (boxer) are hyponyms of ‘hond’ (dog). In principle, the triplets can be used to differentiate between these through their mapping to the ontology. Note that the most frequent structure in the database now is a single triplet with the relation ‘+’. This means that the synset is simply labeled by subsumption to a SUMO concept and nothing is stated about the rigidity of the concept represented by the synset. Such more explicit mappings need to be made in a future extension of the database.

Another case of mixed rigid and non-rigid hyponyms are words for *water*. In the Dutch wordnet there are over 40 words that can be used to refer to water in specific circumstances or with specific attributes. Water is in SUMO a CompoundSubstance just as other molecules. We can thus expect that the synset of *water* in Dutch matches directly to Water in SUMO, just as *zand* matches to Sand. However, *water* has 3 major meanings in the Dutch wordnet: water as liquid, water as a chemical element and a water area, while there are only two concepts in



SUMO: Water as the CompoundSubstance and a WaterArea. In SUMO there is no concept for water in its liquid form, even though this is the most common concept for most people. Most of the hyponyms of *water* in the Dutch Wordnet are linked to the liquid. To properly map them to the ontology, we thus first must map water as a liquid. This can be done by assigning the Attribute Liquid to the concept of Water as a CompoundSubstance. A SUMO axiom for this is:

```
(and (exists ?L ?W)
      (instance, ?W, Water)
      (instance, ?L Liquid)
      (attribute, ?L, ?W) )
```

In the Cornetto database, this complex KIF expression is represented by the simpler relation triplets:

```
(instance, 0, Water)(instance, 1, Liquid) (attribute, 1, 0)
```

We complete this section with a few concrete examples of the structures that can thus be found in the database. The first synset of example (4 below has an inherited SUMO mapping; see 'dwn10_pwn16_pwn20_mapping' after 'name'. The status can be 'true' or 'false' but has not always been set manually. The relation name is here '+', meaning that this synset is a subordinate of 'Artifact'. In this case, the 'arg1' remains empty and thus co-indexes with whatever the synset is used to refer to.

The second synset shows a manual mapping to the ontology. The editor is shown after 'name', the relation `_name` is 'part' and the term in 'arg2' is PlantBranch, meaning that this synset denotes something that is a part of PlantBranch. Note that it is possible that a synset has more than one mapping to the ontology, possibly through multiple equivalence relations to PWN2.0.

In the last last synset of example (4) shows a complex mapping to SUMO was created manually by means of triplets; the 'arg1' is filled with '0', '1' and '2' and the 'arg2' is filled with either terms or '0', '1' and '2' again to define the relation between these terms. For more information on these triplets, see Cornetto Deliverable D03 'Top-level ontology, relation constraints and assignments' for more details and a further discussion.



Note that if a synset has four or more equivalence relations and the synset or its equivalence relation was not checked manually, all the SUMO terms have been deleted. If there are too many equivalence relations, we assume that the mapping to English wordnet is not reliable enough for an automatic import. In those cases, it is better to derive the SUMO label from the hyperonym of the synset in Cornetto.

example (4)

```
(d_n-30184, pop:1)
<sumo_relations>
<ont_relation status="false" name="dwn10_pwn16_pwn20_mapping" negative="false" relation_name="+"
arg1="" arg2="Artifact"/>

(n_n-523427, knoest:2)
<sumo_relations>
<ont_relation status="true" name="piek" negative="false" relation_name="part" arg1="" arg2="PlantBranch"/>

(d_n-20059, theewater:1)
<sumo_relations>
<ont_relation status="false" name="piek" negative="false" relation_name="instance" arg1="0" arg2="Water"/>
<ont_relation status="false" name="piek" negative="false" relation_name="instance" arg1="1" arg2="Tea"/>
<ont_relation status="false" name="piek" negative="false" relation_name="instance" arg1="2"
arg2="Making"/>
<ont_relation status="false" name="piek" negative="false" relation_name="resource" arg1="0" arg2="2"/>
<ont_relation status="false" name="piek" negative="false" relation_name="result" arg1="1" arg2="2"/>
<ont_relation status="false" name="roxane" negative="false" relation_name="component" arg1="0" arg2="1"/>
```

5.1.1.6 Wordnet domains

The domain labels from WordNet Domains (Magnini and Cavaglià 2002) were also imported through the automatically-derived equivalence relations to the English wordnet (PWN1.6 to PWN 2.0). If there were four equivalence relations or more and the synset was not checked manually, the domains were deleted and replaced with a translation of the original DWN labeling to a WordNet domain. For this purpose the domain labels from VLIS were manually aligned with the domain labels in WordNet domains. Each synset can have one or more domain labels like ‘factotum’, ‘music’ or ‘botany’. The set of possible domains is fixed and consists of the 159 domain labels. An overview of the labels is given in Table 6 in section 3.



The labels in WordNet Domains come from the Dewey decimal system and form a hierarchy of 4 levels, which is shown in the Appendix. Higher domain labels are thus implied by lower labels.

Depending on whether the domain mapping was inherited by the equivalence relations, translated from the original DWN domains or created manually or not, the XML of the domain relations may vary slightly. The first example below shows two automatically derived domains: ‘building_industry’ and ‘town_planning’. The second example was created manually; the status is set to true. Note that a status ‘false’ in the domain labels does not imply necessarily that the labeling was not checked. For the status of the labeling the AQL of the synset-LU mapping can be used as an indication or the list of synsets in the appendix with many equivalence relations that are checked manually.

example (5)

```
(d_n-11704, huis:1)
<wn_domains>
<dom_relation status="false" name="dwn10_pwn15_pwn16_mapping" term="building_industry
town_planning"/>

(d_n-26965, auto:1)
<wn_domains>
<dom_relation status="true" name="roxane" term="transport"/>
```

5.1.2 Verbs

The Cornetto verbs follow the same organizing principles and XML structure as the noun synsets. In the following sections only those differences from the nouns will be further described. If there are no comments there will be a redirection to one of the noun paragraphs above.

5.1.2.1 ID Number

Each synset verb ID number (c_sy_id) starts with a ‘d_’ (the synset originates from the original Dutch Wordnet), or a ‘c_’ (the synset is new and created in the Cornetto project) followed by a ‘v-’ for verb as part of speech. In the case of verbs, the POS is differentiated



into: VERB_TRANSITIVE, VERB_INTRANSITIVE, VERB_TRANS_INTRANS, VERB_IMPERSONAL or VERB_REFLEXIVE.

In the next example, the first ID comes from the original DWN, the second was created in the Cornetto project. Note that there is no such POS label in new synsets.

example (6)

(d_v-2324, lopen:2)

```
<cdb_synset posSpecific="VERB_INTRANSITIVE" c_sy_id="d_v-2324" d_synset_id="d_v-2324">
```

(c_602, lopen:3)

```
<cdb_synset c_sy_id="c_602" comment="">
```

5.1.2.2 Synonyms

» See section 5.1.1.2.

5.1.2.3 Internal relations

» See section 5.1.2.3

5.1.2.4 Wordnet Equivalence relations

» See section 5.1.2.4

5.1.2.5 Ontology

» See section 5.1.1.5

5.1.2.6 Wordnet domains

» See section 5.1.1.6

5.1.3 Adjectives

In general the Cornetto adjectives follow the same organizing principles and XML structure as the noun en verb synsets. In the following sections only the features that differ from the nouns will be further described. If there are no comments there will be a reference to one of the noun paragraphs above.



5.1.3.1 ID Number

Each synset adjective ID number (c_sy_id) starts with a 'd_' (the synset originates from the original Dutch Wordnet), a 'n_' (the synset originates from Vlis) or a 'c_' (the synset is new and created in the Cornetto project) followed by an 'a-' for adjective as part of speech. In the next example, the first ID comes from the original DWN, the second from Vlis and the third and fourth were created in the Cornetto project.

example (7)

(d_a-9413, mooi:3)

```
<cdb_synset posSpecific="ADJECTIVE" c_sy_id="d_a-9413" comment="" d_synset_id="d_a-9413">
```

(n_a-515377, mooi:1)

```
cdb_synset posSpecific="ADJECTIVE" c_sy_id="n_a-515377" d_synset_id="n_a-515377">
```

(c_273, lelijk:2)

```
<cdb_synset c_sy_id="c_273" comment="">
```

(c_178, mooi:2)

```
<cdb_synset posSpecific="ADJECTIVE" c_sy_id="c_178" comment="" d_synset_id="d_a-9557">
```

Note that, if a synset is newly created in Cornetto, like the last synset of mooi:2, the 'd_synset_id' points back to the original DWN synset (d_a-9557) from which this new synset is created.

5.1.3.2 Synonyms

» See section 5.1.1.2

5.1.3.3 Internal relations

All adjective synsets are connected to each other by at least one semantic relation, but different from the noun and verb synsets, the adjectives do not necessarily have a hyperonym relation. Internal relations that are most common in adjective synsets are: HAS_HYPERONYM, HAS_XPOS_HYPERONYM, NEAR_SYNONYM, XPOS_NEAR_SYNONYM, NEAR_ANTONYM, STATE_OF and IS_CAUSED_BY.

5.1.3.4 Wordnet equivalence relations

» See section 5.1.1.4



5.1.3.5 Ontology

The adjective synset Princeton Wordnet-to-SUMO mappings are - in contrast with the nouns and verb mappings - not yet corrected manually. This has two major implications: (1) the number of incorrect default mappings is high (2) it has not yet been checked if the ontology is complete. For the mapping of the most frequent Dutch Cornetto adjectives we checked and revised SUMO in order to achieve the coverage of all concepts that are expressed by general language adjectives (see: I.Maks, P. Vossen, R. Segers and H. van der Vliet (2008)).

The new and existing ontology terms that are used in the adjective synsets are listed in the following table.

Table 16: SUMO terms used for Cornetto Adjectives

SUMO and/or Cornetto Class	Description ⁸	Cornetto Examples
Aesthetic Attribute	Attribute used to express someone's opinion about the (lack of) beauty of something or someone	synset n_a-524142- prachtig:1, adembenemend:1 synset d_a-9381- lelijk:1
AgeAttribute	Attribute that characterize the age of someone or somebody	synset d_a-9357- klein:3, jong:1 synset n_a-520979- groen:2, onervaren:1
Animacy Attribute	see SUMO ⁹	synset d_a-9236- dood:1
Appearance Attribute	Attribute that characterizes someone's appearance	synset n_a-517264- snoezig:1, lief:2 synset n_a-533321- vlot:3, sportief:4
Behaviour Attribute	Attribute that characterizes someone's behaviour. ¹⁰	synset n_a-533914- ongedwongen:1, vrij:4, vrijmoedig:1, frank:1, onbevangen:1 synset n_a-517506- lomp:2, plomp, onbehouwen:1, boers:1
Bodily Attribute	Attribute that characterizes someone's body	synset n_a-503569- kaal:1, onbehaard:2
Cognitive Attribute	Attribute that characterizes someone's cognitive abilities	synset c_579- dommig:1, onbenullig:1, oenig:1 synset n_a-527945- slim:1, pienter:1, snugger:1, kien:1

⁸ see SUMO implies that the term refers a original SUMO class; otherwise the term refers to a new ontology class used and defined within Cornetto only

⁹ <http://www.ontologyportal.org/>

¹⁰ In Cornetto, the BehaviourAttributes focus on temporarily character features often evoked by other people's behaviours whereas TraitAttributes focus on more permanent features. However, in many cases this is a rather arbitrary distinction.



		synset n_a-516876- leeg:2, inhoudsloos:1,
ColorAttribute	see SUMO	synset n_a-508890- hel:1, fel:3 synset n_a-509684- geel:1
Consciousness Attribute	see SUMO	synset d_a-9540- duf:1, suf:1 synset n_a-503988- bewust:2
Compositional Attribute	Attribute that characterizes material and refers to its composition	synset n_a-511289 - gouden:1 synset n_a-532927- vet:1, vetrijk:1
Consistency Attribute	Attribute that characterizes material and refers to its consistence	synset d_a-9620 - zacht:1 synset n_a-54423 – week:1
Constitution Attribute	Attribute that characterizes someone's health	synset c_254: slap:2, zwak:1, krachteloos:1, zwakjes:1, slapjes:1
EmotionalState	see SUMO	synset c_254- boos:1, kwaad:1, nijdig:1
LengthAttribute	Attribute that characterizes the length of someone or something	synset d_a-9377- lang:1 synset c_166- klein:2
Negative Evaluation Attribute	Attributes used to express someone's negative opinion	synset c_573- dom:2, duf:2, eentonig:1, droog:4, monotoon:2, slaapverwekkend:1, saai:1, stom:3, vervelend:2 synset d_a-9431- naar:1, akelig:1, onprettig:1, onaangenaam:1, onplezierig:1
Normative Attribute	see SUMO	synset c_80- normaal:1, gewoon:1, regulier:1 synset d_a-9306- juist:1, aangewezen:1
Olfactory Attribute	see SUMO	synset n_a-512200- zalig:1, heerlijk:1, goddelijk:1 synset n_a-533034- vies:1, onsmakelijk:1, smerig:3
PhysicalState	see SUMO	synset n_a-516764- vloeibaar
Positional Attribute	see SUMO	synset d_a-9391- los:1 synset n_a-532245 - ver:1
Positive Evaluation Attribute	Attributes used to express someone's positive opinion	synset c_644- aangenaam:1, aardig:2, fijn:1, heerlijk:2, lekker:3, leuk:1, prettig:1 synset d_a-9323- positief:1, gunstig:2, bevorderlijk:1, favorabel:1
RigidityAttribute	Attribute that refers to the ability of material to stretch or bend	synset n_a-520774- stug:1, stijf:1
Saturation Attribute	see SUMO	synset d_a-9587- klef:1 synset n_a-531691- droog:2, uitgedroogd:1, schraal:1
SizeAttribute	Attribute that characterizes material and refers to its size	synset d_a-9417- smal:1, nauw:1, eng:1 synset n_a-510819- enorm:2, overgroot:1, gigantisch:2
SocialRole	see SUMO	synset n_a-507680- druk:1, bezet:1, geoccupeerd:1- synset d_n-41713- rijk:1, kapitaalkrachtig:1, egoed:1, gefortuneerd:1 synset d_a-9146- edel:3, adellijk:1, aristocratisch:1



SoundAttribute	see SUMO	synset n_a-519240- monotoon:1, eentonig:2 synset n_a-536057zwaar:4, donker:4, diep:5
Subjective Assessment Attribute	see SUMO	vreselijk:1, ontzettend:1, verschrikkelijk:1
TactileAttribute	The class of properties that are detectable by taction	synset c_639- glad:1, effen:2
TasteAttribute	see SUMO	synset n_a-510413- gepeperd:1, heet:2
TemperatureAttrib ute	Attribute that characterizes material and refers to its temperature	synset n_a-515498- koel:1, frisjes:1, fris:1 synset n_a-535263- zacht:4
TextureAttribute	see SUMO	synset c_567- teen:1, fijn:2 synset n_a-514300- kaal:2, versleten:2
TimeAttribute	Attribute that refers to temporal aspects	synset d_a-9408- nieuwerwets:1, eigentijds:1, modern:2 synset n_a-516735- lang:2, langdurig:1, langlopend:1
TraitAttribute	see SUMO, and see BehaviourAttribute	synset c_175- aardig:1, sympathiek:1, vriendelijk:1; synset c_255- kwaadaardig, boos, boosaardig, kwaad, malicieus
VelocityAttribute	Attribute that characterizes the velocity of someone or something	synset n_a-516764- traag:1, langzaam:1 synset n_a-504345- keihard:2, vliegend:3
VisualAttribute	see SUMO	synset d_a-9248- duidelijk:2, helder:2 synset d_a-9386- licht:2
WeightAttribute	Attribute that characterizes the weight of someone or something	synset n_a-517176- licht:1

IntensifierAttribute	Attribute that intensifies the meaning of the concept it modifies	synset c_681 - enorm:2, razend:2, gigantisch:1, buitenmatig:1, immens:2
----------------------	--	--

5.1.3.6 Domains

» See section 5.1.1.6

5.2 Cornetto Lexical Units

The Cornetto Lexical Unit (LU) database is organized around lexical units: a word-meaning combination that represents a lexical form and a single meaning of this form. Each LU can have information on morphology, syntactics, pragmatics, combinatorics, and example sentences that include collocations and idioms. In the next sections we will discuss the main



characteristics and structures for nouns, verbs and adjectives. For more information on the lexical units, see Documentatie Referentie Bestand Nederlands (RBN, Martin et al 2005).

5.2.1 Nouns

5.2.1.1 Lexical unit ID, and sequence number

Each noun lexical unit has its own unique ID number. In the following three examples, the first ID number originates from the RBN and therefore starts with a 'r' for 'RBN' and is in the case of nouns followed by a 'n' for the part of speech. The second example shows an ID that starts with 'c_', meaning that this LU was created in the Cornetto project. The last example shows a new lexical unit as well, but this ID starts with a 'd_'. This LU was initially generated automatically as a dummy LU in the alignment, but was in this case used and completed by one of the editors. The distinction between an edited and non-edited d_ LU can be inferred from the AQL and from the fact that a dummy, thus unedited LU will never have a resume. Note that the LUs that originate from the RBN will be stated 'true' after is_complete; the new LUs (d_ and c_) are not as fully elaborated and therefore notated as being not complete.

The sequence numbers in all lexical units can be seen as the Cornetto sequence numbers, which we indicate in the following by prefixing the sequence number to the word; '10:beer' is therefore the same LU in this part of Cornetto as it is in the Cornetto synset where it will be notated as 'beer:10'. See for example synset d_n-27579:(crediteur:1/r_n-10141,schuldeiser:1/r_n-33414, beer:10/c_545231)

1:beer

```
<cdb_lu c_seq_nr="1" type="swu" is_complete="true" c_lu_id="r_n-6269">
```

```
<form form-cat="noun" form-spelling="beer"/>
```

10:beer

```
<cdb_lu c_seq_nr="10" type="swu" is_complete="false" comment="" c_lu_id="c_545231">
```

```
<form form-length="" form-cat="noun" form-spelling="beer" form-spelvar=""/>
```

8:bal

```
<cdb_lu c_seq_nr="8" type="swu" is_complete="false" comment=""
```

```
c_lu_id="d_n-20569">
```



```
<form form-length="full" form-cat="noun" form-spelling="bal" form-spelvar=""/>
```

5.2.1.2 Morphology

Each noun lexical unit can have values in the field morphology. The morphology is divided into morphological type, morphological structure and plural form as can be seen in the example below. Other values for morpho-type are 'simpmorph', 'compound', 'compderiv', 'nmorph' and 'zero-derivation'. In the case of compounds and derivations, the structure can be notated like '[boek]<en>kast' (compound) or 'molen[*aar]' (derivation).

```
-<morphology_noun>
  <morpho-type>derivation</morpho-type>
  <morpho-structure>mededelen[*ing]</morpho-structure>
- <morpho-plurforms>
  <morpho-plurform>mededelingen</morpho-plurform>
</morpho-plurforms>
</morphology_noun>
```

5.2.1.3 Syntax

The syntax field for nouns can contain information about gender, article, number, numberconcord and complementation. Each of these can have a range of different values like 'f' or 'm' for gender and 'factive' or 'psmodnoun' for syntactic complementation as can be seen in the following two examples. Note that features will not be represented in the XML if they have no value.

1 :mededeling/r n-23410

```
<syntax_noun>
  <sy-gender>f</sy-gender>
  <sy-article>de</sy-article>
<sy-complementation>
  <sy-comp>factive</sy-comp>
  <sy-comp>prep 'aan'</sy-comp>
  <sy-comp>prep 'omtrent'</sy-comp>
  <sy-comp>fixprep 'over'</sy-comp>
```



```

</sy-complementation>
</syntax_noun>

1: zwermer/r_n-44626
<syntax_noun>
    <sy-gender>m</sy-gender>
    <sy-article>de</sy-article>
<sy-complementation>
    <sy-comp>psmodnoun</sy-comp>
</sy-complementation>
</syntax_noun>

```

See also the RBN Documentation (Martin e.a.(2005)) for the extensive list of all possible syntax values.

5.2.1.4 Semantics

The features for the semantics of nouns are: reference, countability, semantic type, semantic shifts, and resume. Evidently, not all of these features will be applicable in every noun LU. Reference can have the value ‘proper’ for proper nouns and ‘common’ for all nouns with a non-unique reference. Countability can have the value ‘count’, ‘mass’, ‘plurant’, ‘coll’, etcetera. The semantic types are organised in a small taxonomy that consists of 13 terms like ‘concrete’, ‘place’, ‘abstract’, ‘institution’, ‘dynamic’, etcetera. These terms are also used to define the semantic shifts like dynamic ? nondynamic. The label shift is used for lexical units that represent actually represent multiple closely related meanings, where one meaning is derived from the other through a shift relation. For example, *artikel* (*article*) is represented as a single LU with a shift that predicts that besides ‘text’, an *artikel* can also be an Artifact.

Furthermore, each edited LU has a resume that functions as a brief definition. These resumes are collided in the definitions of the synsets in which the LU takes part as a synonym.

An example of the XML for the semantics is shown below:

```

1: politie/r_n-29051
<semantics_noun>
    <sem-reference>common</sem-reference>
    <sem-countability>coll</sem-countability>

```



```

<sem-type>institut</sem-type>
<sem-shift>15:inst>hum</sem-shift>
<sem-subclass>institutie</sem-subclass>
<sem-resume>geheel aan ambtenaren v.d. openbare orde</sem-resume>
</semantics_noun>

```

5.2.1.5 Pragmatics

There are eight features related to pragmatics: origin, style, connotation, social group, subject field, chronology, geography and frequency with each a list of possible features. The noun in the following first example has ‘politics’ as subject field and the ‘general= “true” means that this word is domain specific but also used in common language. The value ‘belg’ states that this word is Flemish and not common in the Netherlands. The LU ‘2:kat’ in the second example has a style and a connotation feature with the values ‘informal’ and ‘offensive’.

Note that pragmatics will not be represented in the XML if there are no values.

```

2:schepen/ r n-32897
<pragmatics>
  <prag-domain general="true" subjectfield="pol"/>
  <prag-geography>belg</prag-geography>
</pragmatics>

2:kat/ r n-19177
<pragmatics>
  <prag-style>informal</prag-style>
  <prag-connotation>offens</prag-connotation>
</pragmatics>

```



5.2.1.6 Examples

The examples have a vast amount of features that will not all be discussed in this documentation. For a complete list of features and values, see the RBN documentation (Martin e.a.(2005)). The most important example features that will be discussed here are: example ID, form example, syntax example and semantics example.

The example XML starts with an unique **example ID**; in the case of LUs that start with a 'r_', the example ID is usually a number as can be seen in the first example. In the second example, one of the editors added an example in an already existing LU. In these cases the example ID will start with 'r_n' followed by the ID number of the LU and a sequence number of the example in this LU.

New LUs starting with a 'd_' or 'c_' will have the same structure in their example ID as can be seen in example three and four.

1:aanrecht/r_n-2297

<example r_ex_id="36539">

<canonicalform>een stenen aanrecht</canonicalform>

<example r_ex_id="36540">

<canonicalform>aan het aanrecht (staan)</canonicalform>

3:topper/r_n-38111

<example r_ex_id="r_n-38111-1">

<canonicalform>De professor was een topper in zijn vakgebied</canonicalform>

6:kaart/d_n-129479

<example r_ex_id="d_n-129479-1">

<textualform>Ik heb een paar kaarten over voor de Rolling Stones</textualform>

2:pulp/c_545948

<example r_ex_id="c_545948-1">

<canonicalform>Papier gemaakt van pulp is van veel mindere kwaliteit dan lompenpapier</canonicalform>

The feature form example contains the string of the actual example. Generally all examples are notated after 'canonical form'; the 'textual form' is used for contextualised examples. For each form, the category can specify the structure of the example. In the first example shows



only a canonical form in a np structure; the second example has a canonical form in a pp structure and a contextualised form with a 's' structure (sentence).

Note that most examples in new LUs with an ID number starting with a 'd_' will have an empty canonical form; in these cases only the textual form and the text-category was used. (see example three)

1:aanrecht/r n-2297

```
<form_example>
  <canonicalform>een stenen aanrecht</canonicalform>
  <category>np</category>
</form_example>
```

1:trap/r n-38330

```
<form_example>
  <canonicalform>op een trap</canonicalform>
  <textualform>zij woont op een trap met vrijgezelle vrouwen </textualform>
  <category>pp</category>
  <text-category>s</text-category>
</form_example>
```

8:zaak/d n-341931

```
<form_example>
  <category/>
  <canonicalform/>
  <textualform>Hij is vandaag op de zaak</textualform>
  <text-category>s</text-category>
</form_example>
```

The examples can be divided in free examples that intend to illustrate the meaning of the LU, and fixed examples like idioms, lexical collocations, grammatical collocations, proverbs and slogans.

The feature 'syntax example' specifies if the syntactic type of example is free or fixed. For fixed examples the syntactic subtype specifies if the example is an idiom, lexical collocation etc. The first example shows a free example, the second and third are fixed.



1:laag/r n-21170

(canonicalform: **de beschaving zit er maar met een dun laagje op**)

<syntax_example>

<sy-type>**free**</sy-type>

</syntax_example>

2:roos/r n-32058

(canonicalform: **een schot in de roos**)

<syntax_example>

<sy-type>**fixed**</sy-type>

<sy-subtype>**idiom**</sy-subtype>

</syntax_example>

1:zaak/ r n-43717

(canonicalform: **dat is mijn zaak niet**)

<syntax_example>

<sy-type>**fixed**</sy-type>

<sy-subtype>**pragm**</sy-subtype>

</syntax_example>

Note that some of the features in the syntax example like ‘combinationword’ and ‘cat combinationword’ are not discussed here. See the RBN documentation (Martin e.a.(2005)) for all specifications.

The feature semantics_example can be used for the meaning of the example, the complementation in grammatical collocations and for the notation of the Melcuk lexical function in lexical collocations. This feature is used mostly in fixed examples. The first example below shows an idiom with a meaning description; the second one is a lexical collocation and has a ‘magnus’ lexical function for the collocator. The third example shows a combination of both meaning description, a lexical function and a specification of the collocator in a lexical collocation. The last example is grammatical collocation; the ‘hum/inst’ means that this collocation is necessarily followed by nouns of the semantic type ‘human’ and ‘institute’.

2:roos/r n-32058



(canonicalform: **een schot in de roos**)

<semantics_example>

<sem-meaningdescription>**precies goed**</sem-meaningdescription>

</semantics_example>

1:applaus/r_n-3766

(canonicalform:**een daverend/donderend applaus**)

<semantics_example>

<sem-lc-collocator>**magnus**</sem-lc-collocator>

</semantics_example>

2:zaak/r_n-43718

(canonicalform>**een zaak drijven/runnen**)

<semantics_example>

<sem-meaningdescription>**een handelszaak hebben**</sem-meaningdescription>

<sem-lc-collocator>**oper1**</sem-lc-collocator>

<sem-spec-collocator>**caretake**</sem-spec-collocator>

</semantics_example>

1:opdracht/r_n-26532

(canonicalform>**een opdracht aan/voor [de aannemer]**)

<semantics_example>

<sem-gc-compl>**hum/inst**</sem-gc-compl>

</semantics_example>

5.2.2 Verbs

The Cornetto verb LUs have the same main features and XML structure as the nouns; the main difference are the different sub-features and their values. In the following sections only those differences from the nouns will be further discussed. If there are no comments there will be a redirection to one of the noun paragraphs above.

5.2.2.1 Lexical unit ID, and sequence number

The verb ID numbers and the sequence numbers share the same characteristics as the nouns, the main difference however is that the ‘n’ for noun in the ID number is replaced for a ‘v’ for verb.

» see section 5.2.1.1



5.2.2.2 Morphology

The morphology can be divided into eight sub-features: morphotype, morphostructure, conjugation type, conjugation form, mode, tense, number and person. These features will be discussed by means of the next example. The verb *aantrekken* is ‘phrasal’; this means that the particle [*aan*] can be separated. Other possible values for morphotype are ‘compound’ (*ademhalen*), ‘derivation’ (*dineren*) and ‘simpmorph’ (*lachen*). In flex-conjugation the conjugation type is ‘strong’ stating that the conjugation is irregular. Other values are ‘regular’, ‘mixeda’, ‘mixedb’ and ‘other’. In all irregular cases, there will be a specification of the conjugation as can be seen in ‘pasttense’ and ‘pastpart’. The feature ‘conjugation’ is followed by mode, tense, number and person; these all have the default value. Note that other values are very rare in this features. See the RBN Documentation (Martin e.a.(2005)) for all possible values in morphology.

```
1:aantrekken/r_v-277
<morphology_verb>
  <morpho-type>phrasal</morpho-type>
  <morpho-structure>[aan]trekken</morpho-structure>
<flex-conjugation>
  <flex-conjugationtype>strong</flex-conjugationtype>
  <flex-pasttense>trok aan</flex-pasttense>
  <flex-pastpart>aangetrokken</flex-pastpart>
</flex-conjugation>
  <flex-mode>inf</flex-mode>
  <flex-tense>ntense</flex-tense>
  <flex-number>nnumber</flex-number>
  <flex-person>nperson</flex-person>
</morphology_verb>
```

5.2.2.3 Syntax

Syntax is divided into nine sub-features: class, peraux, valency, separation, complementation, prepcompl (preposition complementation), subject, reflexivity, transitivity. The next example will explain these features. The feature sy-trans is used to specify if the verb is transitive or not. The ‘sch’ in sy-separ means that the particle ‘aan’ can be separated from the head verb. Note that the same was stated in the morpho-type ‘phrasal’ that was discussed in paragraph



5.2.2.2. The sy-class can have three values: main, auxiliary and copula. The sy-peraux is used for the perfective auxiliary that is combined with this verb. In this case it is 'h' voor 'hebben'; other values are 'z' (zijn) and 'hz' ('hebben' and 'zijn'). The sy-valency is used for the number of semantic slots (arguments) the verb has. Possible values are 'mono', 'di' and 'tri'. Note that this classification is different from the more classical transitive/intransitive feature. For example, the verb 1:wonen (r_v-10183) is intransitive and has the value 'di' in valency. The sy-reflexiv is for the reflexivity of the verb, in this case it is not reflexive. Other features are 'oblrefl' (obligatory reflexive) and 'optrefl' (optional reflexive).

The sy-subject can have two values: 'pers' for (personal) and impers (impersonal). The latter is used for verbs that have an impersonal subject like 'het regent/sneeuwt/hagelt' (it's raining/snowing/hailing). The sy-complementation can have a vast amount of values that are not discussed here; in this case the verb needs to be followed by a NP. See the RBN Documentation (Martin e.a.(2005)) for all possible values.



1:aantrekken/ r v-277

```
<syntax_verb>
  <sy-trans>tr</sy-trans>
  <sy-separ>sch</sy-separ>
  <sy-class>main</sy-class>
  <sy-peraux>h</sy-peraux>
  <sy-valency>di</sy-valency>
  <sy-reflexiv>nrefl</sy-reflexiv>
  <sy-subject>pers</sy-subject>
<sy-complementation>
  <sy-comp>np</sy-comp>
</sy-complementation>
</morphology_verb>
```

5.2.2.4 Semantics

The semantics of the verbs is divided in three parts in theXML: semantics_verb, sem-hyperonyms and sem-synonyms. The ‘semantics_verb’ has three main features: semantic type, resume, and caseframe as is shown in the following example. The semantic type of this LU is ‘action’; other possible values are ‘process’ and ‘state’. The sem-caseframe is an action2. For more information on the caseframes, see RBN Documentation (Martin e.a. (2005)). The semantic part ends with a resume that gives a short impression of the meaning of this LU.

2:opschieten/ r v-5748

```
<semantics_verb>
<sem-type>action</sem-type>
  <sem-caseframe>
    <caseframe>action2</caseframe>
  <args>
    <arg>
      <caserole>agent</caserole>
      <selrestrole>agentanimate</selrestrole>
    <synset_list/>
  </arg>
```



```

    <arg>
    <caserole>theme</caserole>
    <selrestrole>themenselres</selrestrole>
    <synset_list/>
  </arg>
</args>
  </sem-caseframe>
<sem-resume>snel gaan</sem-resume>
</semantics_verb>

```

5.2.2.5 Pragmatics

» see section 5.2.1.5

5.2.2.6 Examples

» see section 5.2.1.6

5.2.3 Adjectives

5.2.3.1 Lexical unit ID, and sequence number

The adjective ID numbers and the sequence numbers share the same characteristics as the nouns and verbs; the main difference however is that the ‘n’ for noun in the ID number is replaced for a ‘a’ for adjective.

» see section 5.2.1.1

5.2.3.2 Morphology

The morphology for adjectives is divided into morphological type, morphological structure, flectional type, declinability and comparison. The morphological type of the LU below is ‘derivation’ and the morphological base for this adjective is the verb ‘dragen’. This adjective also has a flectional type; in most cases it has a default value and therefore it will not be represented in the XML, but here it has a value ‘papa’. This means that this form is identical to the past participle of the verb ‘dragen’. Other possible values for flectional type are ‘prespa’ (form is similar to the present participle), ‘pseudoprespa’ and ‘pseudopapa’ (used for adjectives that morphologically look like a past participle or present participle but without a base verb). The feature declinability can have two values; ‘declin’ (default value) and



‘nondeclin’. Finally, comparison has the value ‘periphras’ in this example, meaning that comparative and superlative are formed by means of a periphrase. Other possible values are ‘regular’ (default), ‘ncomparison’, ‘irregular’ and ‘mixed’. For the last two values there also will be a specification of the comparison and superlative.

1:gedragen/ r a-10852

```
<morphology_adj>
  <morpho-type>derivation</morpho-type>
  <mor-base>dragen</mor-base>
  <mor-flectional-type>papa</mor-flectional-type>
  <morpho-structure>'[ge*]dragen'</morpho-structure>
  <mor-declinability>nondeclin</mor-declinability>
  <mor-comparis>
    <mor-comparison>periphras</mor-comparison>
    <mor-comparative>meer gedragen</mor-comparative>
    <mor-superlative>meest gedragen</mor-superlative>
  </mor-comparis>
</morphology_adj>
```

5.2.3.3 Syntax

The main features in syntax are position, adverbial usage and complementation. The first example below has for position the value ‘ap’, meaning that this LU can be used in an attributive and predicative position. Other possible values are ‘attr’ (only attributive) and ‘pred’ (only predicative). Both examples have the value ‘nonadv’ for advusage (adverbial usage); this means that both LUs can not be used as an adverb. Logically the other possible value is ‘nonadv’. The second example also has a complementation feature in the form of a fixed preposition ‘aan’. Note that there is a vast amount of possible values for the complementation patterns. See the RBN Documentation (Martin e.a.(2005)) for a complete overview.



1:rood/r a-14628

```
<syntax_adj>
  <sy-position>ap</sy-position>
  <sy-advusage>nonadv</sy-advusage>
</syntax_adj>
```

1:schuldig/ r a-14866

```
<syntax_adj>
  <sy-position>ap</sy-position>
  <sy-advusage>nonadv</sy-advusage>
  <sy-complementation>
    <sy-comp>fixprep 'aan'</sy-comp>
  </sy-complementation>
</syntax_adj>
```

5.2.3.4 Semantics

The semantics for adjectives has three major features: semantic type, semantic shift, selection restriction and resume. The example below shows a LU of the semantic type ‘abstract’.

Possible values for semantic type for adjectives are: ‘phyper’ (physical/perceptive attributes), ‘stuff’, ‘colour’, ‘emomen’ (emotional/mental attributes), ‘place’, ‘temp’. Furthermore there also is an semantic shift from abstract to emomen. Finally, the semantics conclude with a resume. (for more information about shifts see also section 5.2.1.4 and the RBN Documentation (Martin e.a.(2005)).

2:dor/ r a-10112

```
<semantics_adj>
  <sem-type>abstract</sem-type>
  <sem-shift>abstract>emomen</sem-shift>
  <sem-selrestrictions>
    <selrestriction>concrother</selrestriction>
    <selrestriction>artefact</selrestriction>
    <selrestriction>nondynamic</selrestriction>
  </sem-selrestrictions>
  <sem-resume>saai</sem-resume>
</semantics_adj>
```



5.2.3.5 Pragmatics

» see section 5.2.1.5

5.2.3.6 Examples

» see section 5.2.1.6

6 References

Guarino, N. and Welty, C., (2002). Identity and subsumption. In: R. Green, C. Bean and S. Myaeng (eds.), *The Semantics of Relationships: an Interdisciplinary Perspective*. Kluwer.

Guarino, N. and Welty, C. (2002). Evaluating Ontological Decisions with OntoClean. *Communications of the ACM*, 45(2), 61-65.

Magnini, B., Cavaglià, G. (2000). Integrating subject field codes into WordNet. *Proceedings of the Second International Conference Language Resources and Evaluation Conference (LREC)*, Athens, Greece, 1413–1418.

Maks, I., P. Vossen, R. Segers and H. van der Vliet (2008). *Encoding Adjectives in the Dutch semantic database Cornetto*. In Proceedings of LREC-2008, Marrakech, Morocco, 2008.)

Martin, W. e.a. (2005) Referentiebestand Nederlands : Documentatie Vrije Universiteit Amsterdam : interne publicatie. www.tst.inl.nl , Productcatalogus/RBN

Niles, I., and Pease, A. (2001) Towards a Standard Upper Ontology. In: Proceedings of FOIS 2001, Ogunquit, Maine, pp. 2-9.

Niles, I. and Pease, A. (2003) Mapping WordNet to the Suggested Upper Merged Ontology. Proceedings of the International Conference on Information and Knowledge Engineering, Las Vegas, Nevada.

Niles, I. and Terry, A. (2004) The MILO: A general-purpose, mid-level ontology. Proceedings of the International Conference on Information and Knowledge Engineering. Las Vegas, Nevada.

Tjong Kim Sang, E. and Hofmann, K. (2007). *Automatic Extraction of Dutch Hypernym-Hyponym Pairs*. In Proceedings of CLIN-2006, Leuven, Belgium, 2007.

Vossen, P. (2006). *Cornetto: Een lexicaal-semantische database voor taaltechnologie*, Dixit Special Issue, Stevin.

Vossen, P., Hofmann, K., de Rijke, M., Tjong Kim Sang, E., and Deschacht, K. (2007). The Cornetto Database: Architecture and User-Scenarios. In DIR 2007, pp. 89-96.



Vossen, P., I.Maks, R. Segers and H. van der Vliet (2008). *Integrating Lexical Units, Synsets, and Ontology in the Cornetto Database*. In Proceedings of LREC-2008, Marrakech, Morocco, 2008.

Vossen, P. (ed.) (2002) EuroWordNet General Document1, Version 3 (Final) , University of Amsterdam, <http://www.vossen.info/docs/2002/EWNGeneral.pdf>

- Cornetto Deliverables

All Cornetto Deliverables can be found at the Cornetto website

- Websites

SUMO www.ontologyportal.com

Cornetto <http://www.let.vu.nl/onderzoek/projectsites/cornetto/>

Wordnet <http://wordnet.princeton.edu/>



7 Appendices

7.1 Appendix A: WordnetDomain Hierarchy

- top
 - ..doctrines
 - ..archaeology
 - ..astrology
 - ..history
 -heraldry
 - ..linguistics
 -grammar
 - ..literature
 -philology
 - ..philosophy
 - ..psychology
 -psychoanalysis
 - ..art
 -dance
 -drawing
 -painting
 -philately
 -music
 -photography
 -plastic_arts
 -jewellery
 -numismatics
 -sculpture
 -theatre
 - ..religion
 -mythology
 -occultism
 -roman_catholic
 -theology
 - ..free_time
 -play
 -betting
 -card
 -chess



-sport
 -badminton
 -baseball
 -basketball
 -cricket
 -football
 -golf
 -rugby
 -soccer
 -table_tennis
 -tennis
 -volleyball
 -cycling
 -skating
 -skiing
 -hockey
 -mountaineering
 -rowing
 -swimming
 -sub
 -diving
 -racing
 -athletics
 -wrestling
 -boxing
 -fencing
 -archery
 -fishing
 -hunting
 -bowling
 -showjumping
- ..applied_science
 -agriculture
 -alimentation
 -gastronomy
 -architecture
 -town_planning
 -building_industry
 -furniture



-computer_science
-engineering
 -mechanics
 -astronautics
 -electrotechnics
 -hydraulics
-medicine
 -dentistry
 -pharmacy
 -psychiatry
 -radiology
 -surgery
-veterinary
 -zootechnics
- ..pure_science
 -astronomy
 -topography
 -biology
 -biochemistry
 -ecology
 -botany
 -zoology
 -entomology
 -anatomy
 -physiology
 -genetics
-chemistry
-earth
 -geology
 -meteorology
 -oceanography
 -paleontology
 -geography
-mathematics
 -geometry
 -statistics
-physics
 -acoustics
 -atomic_physic



-electricity
-electronics
-gas
-optics
- ..social_science
-anthropology
-ethnology
-folklore
- ..body_care
- ..military
- ..pedagogy
-school
-university
- ..publishing
- ..sociology
- ..telecommunication
-cinema
-post
-radio
-telegraphy
-telephony
-tv
- ..artisanship
- ..commerce
- ..industry
-textiles
- ..transport
-aeronautic
-auto
-merchant_navy
-railway
- ..economy
-banking
-book_keeping
-enterprise
-exchange
-insurance
-money
-tax



- ..administration
- ..law
- ..politics
 -diplomacy
- ..tourism
- ..fashion
- ..sexuality
- ..factotum
 -number
 -color
 -time_period
 -person
 -quality
 -metrology
 -state



7.2 Appendix B: Manually Aligned Words

a/NOUN	achtertuint/NOUN	afsplitsing/NOUN	ambtenarengerecht/NOUN
aal/NOUN	achteruitgang/NOUN	afpraakje/NOUN	ambtswoning/NOUN
aanbieding/NOUN	achterwerk/NOUN	afstapje/NOUN	amendement/NOUN
aandeelhouder/NOUN	achterzijde/NOUN	afstraffing/NOUN	amerikaan/NOUN
aandrang/NOUN	acquisitie/NOUN	afstudeerscriptie/NOUN	amfitheater/NOUN
aaneenschakeling/NOUN	act/NOUN	af titeling/NOUN	amplitude/NOUN
aanleg/NOUN	acteertalent/NOUN	afvaardiging/NOUN	analist/NOUN
aanloop/NOUN	actie/NOUN	afvalberg/NOUN	analyse/NOUN
aanmaning/NOUN	activiteit/NOUN	afvalhoop/NOUN	ananas/NOUN
aanslag/NOUN	adaptatie/NOUN	afvoer/NOUN	anarchie/NOUN
aanslagbiljet/NOUN	addenda/NOUN	afweging/NOUN	andijvie/NOUN
aansluiting/NOUN	ademhaling/NOUN	afwezigheid/NOUN	anesthesie/NOUN
aanvechting/NOUN	ademtest/NOUN	afwikkeling/NOUN	anijs/NOUN
aanvoer/NOUN	ader/NOUN	afwisseling/NOUN	anker/NOUN
aanwas/NOUN	administratie/NOUN	afzegging/NOUN	annalen/NOUN
aanzien/NOUN	administratiekantoor/NOU	afzet/NOUN	annonce/NOUN
aanzoek/NOUN	N	afzetting/NOUN	annulering/NOUN
aarde/NOUN	admiraliteit/NOUN	agent/NOUN	ansjovis/NOUN
aarzeling/NOUN	adreswijziging/NOUN	agentschap/NOUN	antagonist/NOUN
aas/NOUN	advertentie/NOUN	agressor/NOUN	antenne/NOUN
abattoir/NOUN	advocatenkantoor/NOUN	airmail/NOUN	antiekbeurs/NOUN
abdij/NOUN	afbakening/NOUN	akker/NOUN	antiekwinkel/NOUN
abortuskliniek/NOUN	afbetaling/NOUN	akkoord/NOUN	antiquaar/NOUN
ABP/NOUN	afboeking/NOUN	akte/NOUN	antiquair/NOUN
academie/NOUN	afdeling/NOUN	alarmcentrale/NOUN	antiquariaat/NOUN
accent/NOUN	afdeling/NOUN	alcoholgebruik/NOUN	antiquiteit/NOUN
accident/NOUN	afghaan/NOUN	alcoholtest/NOUN	apothek/NOUN
accommodatie/NOUN	afgod/NOUN	alfa/NOUN	apparaat/NOUN
accountantskantoor/NOU	afgraving/NOUN	algengroei/NOUN	appartement/NOUN
N	afkickcentrum/NOUN	alineaa/NOUN	appel/NOUN
accountantsverklaring/NO	afkoeling/NOUN	alkoof/NOUN	apv/NOUN
UN	afleidingsmanoeuvre/NOU	alleenspraak/NOUN	aquamarijn/NOUN
accumulatie/NOUN	N	allemansvriend/NOUN	arabesk/NOUN
acht/NOUN	afnemer/NOUN	alleseter/NOUN	arbeidsbemiddeling/NOU
achterbak/NOUN	afperser/NOUN	alliantie/NOUN	N
achterban/NOUN	afrit/NOUN	allocatie/NOUN	arbeidsbureau/NOUN
achterblijver/NOUN	afscheid/NOUN	alt/NOUN	arbeidscontract/NOUN
achterdek/NOUN	af scheiding/NOUN	altaar/NOUN	arbeidsduurverkorting/NO
achtergrond/NOUN	af scheidscadeau/NOUN	alternatief/NOUN	UN
achtergrondkoor/NOUN	af scheidswoord/NOUN	amalgama/NOUN	arbeidsovereenkomst/NO
achterkamer/NOUN	afschrijving/NOUN	amandelontsteking/NOUN	UN
achterkant/NOUN	afschrikking/NOUN	amazone/NOUN	arbeidstherapie/NOUN
achterschip/NOUN	afslag/NOUN	ambassade/NOUN	arbeidstijdverkorting/NOU
achterste/NOUN	afsluiter/NOUN	amber/NOUN	N
achtersteven/NOUN	afsluiting/NOUN	ambivalentie/NOUN	arbeidsverdeling/NOUN



arbitrage/NOUN	automatenhal/NOUN	balk/NOUN	bedenking/NOUN
arbitragehof/NOUN	automatiek/NOUN	balkenbrij/NOUN	bedenktijd/NOUN
arboretum/NOUN	automatisering/NOUN	balkon/NOUN	bediening/NOUN
archetype/NOUN	autoriteit/NOUN	balkondeur/NOUN	beding/NOUN
archief/NOUN	autosloperij/NOUN	ballade/NOUN	bedoeling/NOUN
archiefbeelden/NOUN	autosnelweg/NOUN	ballentent/NOUN	bedoening/NOUN
archiefkast/NOUN	autostrade/NOUN	ballon/NOUN	bedrading/NOUN
architectenbureau/NOUN	autoverkeer/NOUN	bami/NOUN	bedrijf/NOUN
architectuur/NOUN	autoweg/NOUN	ban/NOUN	bedrijfspan/NOUN
arena/NOUN	avenue/NOUN	band/NOUN	bedrog/NOUN
argumentatie/NOUN	avondblad/NOUN	bandbreedte/NOUN	bedstee/NOUN
aria/NOUN	avondeten/NOUN	bandenspanning/NOUN	beeld/NOUN
ark/NOUN	avondmaal/NOUN	bandje/NOUN	beeldcultuur/NOUN
arm/NOUN	avondmaaltijd/NOUN	bandopname/NOUN	been/NOUN
armenzorg/NOUN	avondschemering/NOUN	banenmarkt/NOUN	beenbreuk/NOUN
arrest/NOUN	avondwinkel/NOUN	bank/NOUN	beenfractuur/NOUN
arrondissementsrechtbank/ NOUN	avonturenfilm/NOUN	banketbakker/NOUN	beer/NOUN
arsenaal/NOUN	avonturenroman/NOUN	bankier/NOUN	beet/NOUN
articulatie/NOUN	avontuurtje/NOUN	bankroof/NOUN	bef/NOUN
artikel/NOUN	b/NOUN	bantamgewicht/NOUN	begaafdheid/NOUN
artisjok/NOUN	baal/NOUN	banvloek/NOUN	begeleider/NOUN
artotheek/NOUN	baan/NOUN	bar/NOUN	begeleiding/NOUN
as/NOUN	baanvak/NOUN	barbaarsheid/NOUN	beginkapitaal/NOUN
asbak/NOUN	baar/NOUN	barbecue/NOUN	beginpunt/NOUN
asiel/NOUN	baard/NOUN	barenswee/NOUN	beginselverklaring/NOUN
assemblage/NOUN	baas/NOUN	bariton/NOUN	begravenisstoet/NOUN
assimilatie/NOUN	babbel/NOUN	barrière/NOUN	begrip/NOUN
assistentie/NOUN	babbeltje/NOUN	barst/NOUN	begroeting/NOUN
assortiment/NOUN	backgammon/NOUN	bas/NOUN	begroting/NOUN
assurantiekantoor/NOUN	bacteriologie/NOUN	baseball/NOUN	behandeling/NOUN
ast/NOUN	bad/NOUN	baseline/NOUN	behandelkamer/NOUN
atelier/NOUN	badge/NOUN	basis/NOUN	beharing/NOUN
Atlas/NOUN	badhuis/NOUN	basispakket/NOUN	beheer/NOUN
atletiekbaan/NOUN	badkuip/NOUN	basisschool/NOUN	behuizing/NOUN
atmosfeer/NOUN	badstof/NOUN	basisvorming/NOUN	bejaardencentrum/NOUN
atoomcentrale/NOUN	badzout/NOUN	basketbal/NOUN	bejaardenhuis/NOUN
atrium/NOUN	bagage/NOUN	bast/NOUN	bejaardenoord/NOUN
attentie/NOUN	bagagerek/NOUN	bastaard/NOUN	bejaardentehuis/NOUN
attractie/NOUN	bak/NOUN	bastion/NOUN	bejaardenwerk/NOUN
attribuut/NOUN	bakker/NOUN	batterij/NOUN	bejaardenwoning/NOUN
aula/NOUN	bakkerij/NOUN	bazaar/NOUN	bejaardenzorg/NOUN
aureool/NOUN	bakkie/NOUN	beademing/NOUN	bek/NOUN
auto/NOUN	baklucht/NOUN	bebossing/NOUN	bekendmaking/NOUN
autobaan/NOUN	baksteen/NOUN	bebouwing/NOUN	bekerfinale/NOUN
autobus/NOUN	bal/NOUN	bed/NOUN	bekken/NOUN
autohandel/NOUN	balans/NOUN	bedankje/NOUN	bekleding/NOUN
autohandelaar/NOUN	balcontrole/NOUN	bedding/NOUN	bekrachtiging/NOUN
automaat/NOUN	balg/NOUN	bede/NOUN	bekroning/NOUN
	balie/NOUN	bedeling/NOUN	bel/NOUN



belasting/NOUN	beroep/NOUN	bijklink/NOUN	boeking/NOUN
belastingaangifte/NOUN	beroepsonderwijs/NOUN	bijlage/NOUN	boekwinkel/NOUN
belastingaanslag/NOUN	beroepsopleiding/NOUN	bijles/NOUN	boel/NOUN
belastingkantoor/NOUN	beroving/NOUN	bijproduct/NOUN	Boer/NOUN
belediging/NOUN	bes/NOUN	bijstand/NOUN	boerderij/NOUN
beleefdheid/NOUN	beschermengel/NOUN	bijstandswet/NOUN	boerenbedrijf/NOUN
beleg/NOUN	beschermers/NOUN	bijt/NOUN	boerenkool/NOUN
belegger/NOUN	beschermheilige/NOUN	biljart/NOUN	boerin/NOUN
belegging/NOUN	beschuldigde/NOUN	binding/NOUN	boetedoening/NOUN
beleidsplan/NOUN	beslag/NOUN	binnenbaan/NOUN	boetiek/NOUN
belemmering/NOUN	besluit/NOUN	binnenbad/NOUN	boezem/NOUN
belevenis/NOUN	bespreking/NOUN	biopsie/NOUN	bok/NOUN
belfort/NOUN	best/NOUN	bips/NOUN	bol/NOUN
belichting/NOUN	bestaan/NOUN	bisschoppenconferentie/NOUN	bolero/NOUN
belijdenis/NOUN	bestand/NOUN		bolletje/NOUN
beloning/NOUN	besteding/NOUN	bitter/NOUN	bom/NOUN
beloop/NOUN	bestek/NOUN	blaadje/NOUN	bon/NOUN
bemesting/NOUN	bestraffing/NOUN	blaag/NOUN	bond/NOUN
bemiddelingsbureau/NOUN	besturing/NOUN	blaar/NOUN	bondgenoot/NOUN
N	bestuurder/NOUN	blaas/NOUN	bonenstaak/NOUN
bemiddelingspoging/NOUN	betaling/NOUN	blad/NOUN	bonthandel/NOUN
N	betalingsbewijs/NOUN	bladmuziek/NOUN	boodschap/NOUN
bemoediging/NOUN	betoging/NOUN	bladvulling/NOUN	boog/NOUN
bende/NOUN	betoog/NOUN	blaffer/NOUN	boom/NOUN
benedenhuis/NOUN	betovering/NOUN	blijf-van-mijn-	boord/NOUN
benedenverdieping/NOUN	betrekking/NOUN	lijfhuis/NOUN	booreiland/NOUN
benzinepomp/NOUN	beurs/NOUN	blijspel/NOUN	boorplatform/NOUN
benzinstation/NOUN	beursbericht/NOUN	blik/NOUN	boosdoener/NOUN
beoordeling/NOUN	bevestiging/NOUN	bliksemafleider/NOUN	bord/NOUN
bepaling/NOUN	bevrijdingsbeweging/NOUN	blind/NOUN	bordeel/NOUN
beraad/NOUN	N	bloem/NOUN	border/NOUN
beraadslaging/NOUN	bevruchting/NOUN	bloembed/NOUN	borg/NOUN
bereik/NOUN	bewapening/NOUN	bloemenstalletje/NOUN	borst/NOUN
berekening/NOUN	bewerking/NOUN	bloemenwinkel/NOUN	borstel/NOUN
berenmuts/NOUN	bezetter/NOUN	bloemist/NOUN	borstvoeding/NOUN
berg/NOUN	bezetting/NOUN	blok/NOUN	borstwering/NOUN
bergformatie/NOUN	bezwinging/NOUN	bobsleebaan/NOUN	bos/NOUN
berghelling/NOUN	B-film/NOUN	bodem/NOUN	bosbes/NOUN
berghut/NOUN	bibliobus/NOUN	bodemvorming/NOUN	bosbouw/NOUN
berging/NOUN	biechtstoel/NOUN	boedelbeschrijving/NOUN	bosje/NOUN
bergketen/NOUN	bierbrouwer/NOUN	boedelscheiding/NOUN	bot/NOUN
bergland/NOUN	bierbrouwerij/NOUN	boedelveiding/NOUN	boter-kaas-en-
berglandschap/NOUN	bierbuik/NOUN	boef/NOUN	eieren/NOUN
bergmassief/NOUN	bies/NOUN	boegbeeld/NOUN	boterletter/NOUN
bergtop/NOUN	bijbel/NOUN	boek/NOUN	bouclé/NOUN
bergwand/NOUN	bijbelboek/NOUN	boekenbeurs/NOUN	boudoir/NOUN
bergweide/NOUN	bijdrage/NOUN	boekhandel/NOUN	bout/NOUN
berisping/NOUN	bijeenkomst/NOUN	boekhandelaar/NOUN	bouw/NOUN
berm/NOUN	bijkantoor/NOUN	boekhouding/NOUN	bouwblok/NOUN



bouwplan/NOUN	buffet/NOUN	centrale/NOUN	conjunctie/NOUN
bouwput/NOUN	buggy/NOUN	centrum/NOUN	connectie/NOUN
bouwsteen/NOUN	buiging/NOUN	chanson/NOUN	console/NOUN
bovenkant/NOUN	buikkrimp/NOUN	charter/NOUN	constellatie/NOUN
bovenlaag/NOUN	buikloop/NOUN	chemotherapie/NOUN	constitutie/NOUN
bovenverdieping/NOUN	buil/NOUN	chili/NOUN	constructie/NOUN
bovenwoning/NOUN	buis/NOUN	Chinees/NOUN	consumptie/NOUN
bovenzaal/NOUN	buit/NOUN	cilinder/NOUN	contact/NOUN
bovenzijde/NOUN	buiten/NOUN	cinema/NOUN	contingent/NOUN
bowl/NOUN	buitenhaven/NOUN	circuit/NOUN	contract/NOUN
bowling/NOUN	buitenopname/NOUN	circulaire/NOUN	controlepost/NOUN
box/NOUN	bul/NOUN	circulatie/NOUN	convent/NOUN
braam/NOUN	bundel/NOUN	circus/NOUN	conventie/NOUN
braille/NOUN	bungalowpark/NOUN	civilisatie/NOUN	conversie/NOUN
branche/NOUN	burcht/NOUN	classicisme/NOUN	convocatie/NOUN
brandgang/NOUN	bureau/NOUN	climax/NOUN	corps/NOUN
brandstofverbruik/NOUN	burg/NOUN	closet/NOUN	corpus/NOUN
brandweer/NOUN	bus/NOUN	clownerie/NOUN	correctie/NOUN
brandweerkazerne/NOUN	buste/NOUN	club/NOUN	correlatie/NOUN
brandwondencentrum/NO	buurt/NOUN	clubgebouw/NOUN	correspondent/NOUN
UN	buurtcentrum/NOUN	clubhuis/NOUN	correspondentie/NOUN
brasserie/NOUN	buurthuis/NOUN	cockpit/NOUN	corrosie/NOUN
break/NOUN	buurthuiswerk/NOUN	code/NOUN	countertenor/NOUN
breuk/NOUN	buurtwacht/NOUN	college/NOUN	coupe/NOUN
breuklijn/NOUN	buurtwerk/NOUN	collegezaal/NOUN	coupé/NOUN
breukvlak/NOUN	buurtwinkel/NOUN	coma/NOUN	coupure/NOUN
brevet/NOUN	B-weg/NOUN	combinatie/NOUN	couvert/NOUN
brevier/NOUN	c/NOUN	commandant/NOUN	cover/NOUN
brief/NOUN	cabaret/NOUN	commando/NOUN	creatie/NOUN
brievenbus/NOUN	cabine/NOUN	commentaar/NOUN	creativiteit/NOUN
brigadier/NOUN	cacao/NOUN	commissariaat/NOUN	credit/NOUN
brik/NOUN	cachot/NOUN	commissaris/NOUN	crème/NOUN
bril/NOUN	cadans/NOUN	commissie/NOUN	crisis/NOUN
brillenkoker/NOUN	café/NOUN	communicant/NOUN	crisistijd/NOUN
broccoli/NOUN	cafetaria/NOUN	communisme/NOUN	cultuur/NOUN
broeder/NOUN	canon/NOUN	compagnie/NOUN	curie/NOUN
broederschap/NOUN	capitulatie/NOUN	compartiment/NOUN	cursus/NOUN
broeikaseffect/NOUN	capsule/NOUN	competentie/NOUN	cyclus/NOUN
broek/NOUN	carambole/NOUN	compilatie/NOUN	daalder/NOUN
brokstuk/NOUN	carnaval/NOUN	complex/NOUN	dag/NOUN
brompot/NOUN	carport/NOUN	compositie/NOUN	dagboek/NOUN
bron/NOUN	cash-and-carry/NOUN	computerbedrijf/NOUN	dak/NOUN
brood/NOUN	casino/NOUN	computersysteem/NOUN	daling/NOUN
broodjeszaak/NOUN	catechisatie/NOUN	concentratie/NOUN	dam/NOUN
broodmaaltijd/NOUN	cel/NOUN	concessie/NOUN	dame/NOUN
brug/NOUN	celstof/NOUN	conferentietafel/NOUN	dans/NOUN
bruggenhoofd/NOUN	celstraf/NOUN	configuratie/NOUN	dansschool/NOUN
brugleuning/NOUN	cement/NOUN	confiscatie/NOUN	danstheater/NOUN
buffer/NOUN	sensor/NOUN	congregatie/NOUN	das/NOUN



dealer/NOUN	dragonder/NOUN	elektriciteit/NOUN	faculteit/NOUN
decreet/NOUN	drankgebruik/NOUN	elektro-	fakir/NOUN
deel/NOUN	dressuur/NOUN	encefalogram/NOUN	familiehotel/NOUN
deelname/NOUN	dreutel/NOUN	element/NOUN	fanmail/NOUN
defensie/NOUN	driepoot/NOUN	ellips/NOUN	fantasie/NOUN
dek/NOUN	drift/NOUN	embargo/NOUN	farce/NOUN
deken/NOUN	drinkwatervoorziening/NO	employé/NOUN	farmacie/NOUN
dekking/NOUN	UN	energiecentrale/NOUN	fascisme/NOUN
deletie/NOUN	drogist/NOUN	engagement/NOUN	fase/NOUN
deling/NOUN	dromer/NOUN	engel/NOUN	fasering/NOUN
demonstratie/NOUN	droogte/NOUN	engelbewaarder/NOUN	fauna/NOUN
departement/NOUN	drop/NOUN	enkel/NOUN	favoriet/NOUN
depot/NOUN	druif/NOUN	enkelband/NOUN	fazant/NOUN
depressie/NOUN	druk/NOUN	enkeltje/NOUN	feestgedruis/NOUN
derivatie/NOUN	drukker/NOUN	enquête/NOUN	feestmaal/NOUN
derrière/NOUN	drukke/NOUN	ensemble/NOUN	feestprogramma/NOUN
detailhandel/NOUN	dualisme/NOUN	entree/NOUN	feitelijkheid/NOUN
detective/NOUN	dubbel/NOUN	episcopaat/NOUN	felheid/NOUN
dienst/NOUN	dubbeltje/NOUN	epos/NOUN	feliciteit/NOUN
diesel/NOUN	dufheid/NOUN	equipage/NOUN	feminisme/NOUN
dikte/NOUN	duiker/NOUN	erfenis/NOUN	fenomeen/NOUN
dinar/NOUN	duim/NOUN	erkenning/NOUN	fiasco/NOUN
diplomatie/NOUN	duinafslag/NOUN	ernst/NOUN	figurant/NOUN
discant/NOUN	dump/NOUN	ervaring/NOUN	figuur/NOUN
discotheek/NOUN	dwaasheid/NOUN	etage/NOUN	filatelie/NOUN
display/NOUN	dwarsligger/NOUN	eten/NOUN	filiaal/NOUN
dispuut/NOUN	dweil/NOUN	eter/NOUN	film/NOUN
distinctie/NOUN	echo/NOUN	etherpiraat/NOUN	filmopname/NOUN
divisie/NOUN	echografie/NOUN	eucharistie/NOUN	filmproducent/NOUN
doek/NOUN	echoscopie/NOUN	evaluatie/NOUN	filter/NOUN
doktersroman/NOUN	economie/NOUN	evangelie/NOUN	financier/NOUN
dol/NOUN	edict/NOUN	exces/NOUN	firma/NOUN
dollar/NOUN	editie/NOUN	executie/NOUN	firment/NOUN
dominicaan/NOUN	eend/NOUN	exegese/NOUN	fixatie/NOUN
doodsoorzaak/NOUN	eenheid/NOUN	existentie/NOUN	flank/NOUN
doorn/NOUN	eenvoud/NOUN	expansie/NOUN	flap/NOUN
doorsnede/NOUN	eer/NOUN	expediteur/NOUN	flard/NOUN
dop/NOUN	eeuwigheid/NOUN	expeditie/NOUN	flat/NOUN
dorsmachine/NOUN	eeuwwisseling/NOUN	expertise/NOUN	flatgebouw/NOUN
dot/NOUN	effect/NOUN	explicatie/NOUN	fles/NOUN
douanekantoor/NOUN	effectenhandel/NOUN	exporteur/NOUN	flipper/NOUN
doublure/NOUN	EHBO/NOUN	expositie/NOUN	flits/NOUN
douche/NOUN	ei/NOUN	expressie/NOUN	flitslicht/NOUN
draad/NOUN	eikel/NOUN	extrapolatie/NOUN	fluit/NOUN
draaiboek/NOUN	eind/NOUN	eyeliner/NOUN	fok/NOUN
draaiing/NOUN	einddiploma/NOUN	ezel/NOUN	fondant/NOUN
Draak/NOUN	eis/NOUN	f/NOUN	fonds/NOUN
drachme/NOUN	elastiek/NOUN	fabel/NOUN	fonkeling/NOUN
drager/NOUN	elektra/NOUN	faciliteit/NOUN	fonteintje/NOUN



foor/NOUN	gebeuren/NOUN	getuige/NOUN	gries/NOUN
formatie/NOUN	gebied/NOUN	geul/NOUN	groei/NOUN
formule/NOUN	gebiedsuitbreiding/NOUN	geus/NOUN	groen/NOUN
fortuin/NOUN	geboorte/NOUN	gevaar/NOUN	groenstrook/NOUN
forum/NOUN	geboorteaangifte/NOUN	geval/NOUN	groentetuin/NOUN
fotoserie/NOUN	gebrabbel/NOUN	gevoel/NOUN	groep/NOUN
fout/NOUN	gebrek/NOUN	gewas/NOUN	groeve/NOUN
foxtrot/NOUN	gedachte/NOUN	geweld/NOUN	grofheid/NOUN
fractuur/NOUN	gedonder/NOUN	gewelddadigheid/NOUN	grond/NOUN
fragment/NOUN	geel/NOUN	gewest/NOUN	grondregel/NOUN
franchise/NOUN	geest/NOUN	gewicht/NOUN	grondstation/NOUN
frase/NOUN	geesteskind/NOUN	gezag/NOUN	grootheid/NOUN
frequentie/NOUN	gegeven/NOUN	gezant/NOUN	grootmeester/NOUN
frictie/NOUN	geheim/NOUN	gezantschap/NOUN	grootseheid/NOUN
frisbee/NOUN	geheimschrift/NOUN	gezegde/NOUN	gruwel/NOUN
front/NOUN	geheimtaal/NOUN	gezel/NOUN	gummi/NOUN
fruithandel/NOUN	geheimzinnigheid/NOUN	gezicht/NOUN	guts/NOUN
fruitteelt/NOUN	geheugen/NOUN	gier/NOUN	haak/NOUN
functie/NOUN	gehuil/NOUN	gift/NOUN	haakje/NOUN
fundament/NOUN	geit/NOUN	giro/NOUN	haal/NOUN
furie/NOUN	geitenkaas/NOUN	glaasje/NOUN	haan/NOUN
futurisme/NOUN	gek/NOUN	gladheid/NOUN	haar/NOUN
fysiotherapie/NOUN	gekheid/NOUN	glans/NOUN	hak/NOUN
g/NOUN	geknoei/NOUN	glas/NOUN	hals/NOUN
gaanderij/NOUN	gel/NOUN	glazuur/NOUN	ham/NOUN
gaard/NOUN	geleding/NOUN	gleuf/NOUN	hamer/NOUN
gal/NOUN	gelegenheid/NOUN	gloed/NOUN	handel/NOUN
galerie/NOUN	geleiding/NOUN	god/NOUN	handeling/NOUN
galerij/NOUN	geloof/NOUN	godshuis/NOUN	handigheid/NOUN
galg/NOUN	geluidsopname/NOUN	goed/NOUN	handwerk/NOUN
galjoen/NOUN	gemaal/NOUN	goederenvervoer/NOUN	hanenkam/NOUN
galmgat/NOUN	gemak/NOUN	goedheid/NOUN	hap/NOUN
galop/NOUN	gemeenschap/NOUN	goedkeuring/NOUN	hark/NOUN
gamma/NOUN	gemeente/NOUN	golf/NOUN	harmonica/NOUN
gang/NOUN	gemeentearchief/NOUN	gondel/NOUN	harmonika/NOUN
gangmaker/NOUN	gemeentehuis/NOUN	gordel/NOUN	hart/NOUN
gangpad/NOUN	genade/NOUN	goud/NOUN	havenhoofd/NOUN
ganzenbord/NOUN	gendarmerie/NOUN	goudmijn/NOUN	Heer/NOUN
garage/NOUN	generator/NOUN	gouverneur/NOUN	heerlijkheid/NOUN
garde/NOUN	genie/NOUN	graad/NOUN	hei/NOUN
gas/NOUN	genoegen/NOUN	graan/NOUN	hek/NOUN
gasbedrijf/NOUN	gerecht/NOUN	graanoogst/NOUN	hel/NOUN
gasfabriek/NOUN	gericht/NOUN	grafiek/NOUN	held/NOUN
gasrekening/NOUN	gerommel/NOUN	granaat/NOUN	helling/NOUN
gast/NOUN	geschiedenis/NOUN	gratie/NOUN	hemel/NOUN
gastheer/NOUN	gesel/NOUN	grauw/NOUN	herhaling/NOUN
gasthuis/NOUN	geslacht/NOUN	green/NOUN	herinnering/NOUN
gat/NOUN	gestel/NOUN	greep/NOUN	hernhutter/NOUN
gebabbel/NOUN	getal/NOUN	griend/NOUN	herrie/NOUN



hersens/NOUN	interferentie/NOUN	kas/NOUN	klokkenspel/NOUN
herstel/NOUN	intermezzo/NOUN	kast/NOUN	kluif/NOUN
hertshoom/NOUN	interpretatie/NOUN	kastanje/NOUN	knaap/NOUN
hindernis/NOUN	intimiteit/NOUN	kasteel/NOUN	knabbeltje/NOUN
hit/NOUN	intrede/NOUN	kat/NOUN	knak/NOUN
hoek/NOUN	introductie/NOUN	katalysator/NOUN	kneep/NOUN
hoepel/NOUN	invoer/NOUN	katapult/NOUN	knip/NOUN
hof/NOUN	inzet/NOUN	kater/NOUN	knoedel/NOUN
hogeschool/NOUN	inzicht/NOUN	katoen/NOUN	knoeier/NOUN
hok/NOUN	isolatie/NOUN	kattenbak/NOUN	knol/NOUN
hokje/NOUN	jacht/NOUN	kattengat/NOUN	knoop/NOUN
hol/NOUN	jager/NOUN	kb/NOUN	knop/NOUN
honger/NOUN	jeans/NOUN	keelklank/NOUN	knots/NOUN
hoofd/NOUN	jong/NOUN	keet/NOUN	koe/NOUN
hoofdeinde/NOUN	jongen/NOUN	kegel/NOUN	koepel/NOUN
hoogachting/NOUN	jood/NOUN	kei/NOUN	koers/NOUN
hoogte/NOUN	juffertje/NOUN	kenmerk/NOUN	koffie/NOUN
hoop/NOUN	juffrouw/NOUN	kennel/NOUN	kogel/NOUN
hoorn/NOUN	jumbo/NOUN	kerk/NOUN	kolf/NOUN
horde/NOUN	juweel/NOUN	kerkenraad/NOUN	kolk/NOUN
horizon/NOUN	kaak/NOUN	kern/NOUN	kolom/NOUN
hospitium/NOUN	kaars/NOUN	ketting/NOUN	kolonie/NOUN
houding/NOUN	kaart/NOUN	keuken/NOUN	komedie/NOUN
huis/NOUN	kabel/NOUN	kiel/NOUN	komma/NOUN
huls/NOUN	kabinet/NOUN	kiezel/NOUN	konijn/NOUN
idee/NOUN	kabouter/NOUN	kijker/NOUN	koning/NOUN
idoom/NOUN	kade/NOUN	kilogram/NOUN	koningin/NOUN
ijs/NOUN	kader/NOUN	kist/NOUN	kont/NOUN
ijzerwaren/NOUN	kalebas/NOUN	kit/NOUN	konvooi/NOUN
illegaliteit/NOUN	kalender/NOUN	klacht/NOUN	kooi/NOUN
impuls/NOUN	kalf/NOUN	klap/NOUN	kooldioxide/NOUN
inbreng/NOUN	kaliber/NOUN	klapper/NOUN	koor/NOUN
index/NOUN	kalk/NOUN	klas/NOUN	kop/NOUN
infiltratie/NOUN	kam/NOUN	klasse/NOUN	koper/NOUN
ingang/NOUN	kameleon/NOUN	klassieker/NOUN	kopie/NOUN
inhoud/NOUN	kamer/NOUN	klauw/NOUN	koppel/NOUN
inlichting/NOUN	kamp/NOUN	klavier/NOUN	koppeling/NOUN
inrichting/NOUN	kanaal/NOUN	kleed/NOUN	kopstoot/NOUN
inschrijving/NOUN	kandidaat/NOUN	kleintje/NOUN	kopstuk/NOUN
insecteneter/NOUN	kanjer/NOUN	klem/NOUN	korps/NOUN
inslag/NOUN	kanon/NOUN	klep/NOUN	kost/NOUN
installatie/NOUN	kant/NOUN	klepper/NOUN	kot/NOUN
instantie/NOUN	kantlijn/NOUN	klets/NOUN	kousenband/NOUN
instelling/NOUN	kap/NOUN	kletskop/NOUN	kraag/NOUN
instituut/NOUN	kapel/NOUN	kleur/NOUN	kraai/NOUN
instructie/NOUN	kapitaal/NOUN	kleurtje/NOUN	kraak/NOUN
instrument/NOUN	kapittel/NOUN	klis/NOUN	kraal/NOUN
integriteit/NOUN	karakter/NOUN	klit/NOUN	kraamkamer/NOUN
interesse/NOUN	karton/NOUN	klok/NOUN	kraan/NOUN



kracht/NOUN	legaat/NOUN	maagd/NOUN	monster/NOUN
kraker/NOUN	legende/NOUN	maaksel/NOUN	moor/NOUN
kras/NOUN	leger/NOUN	maal/NOUN	mop/NOUN
Kreeft/NOUN	legger/NOUN	maand/NOUN	moraal/NOUN
kreng/NOUN	legioen/NOUN	maat/NOUN	morfologie/NOUN
krent/NOUN	lei/NOUN	maatschappij/NOUN	morgen/NOUN
krib/NOUN	leiding/NOUN	macht/NOUN	motief/NOUN
Krijt/NOUN	lél/NOUN	majoor/NOUN	motivatie/NOUN
krijtstreep/NOUN	lengte/NOUN	man/NOUN	motor/NOUN
kritiek/NOUN	lens/NOUN	mandaat/NOUN	muil/NOUN
kroon/NOUN	les/NOUN	maniak/NOUN	muis/NOUN
krop/NOUN	letterkast/NOUN	maniertje/NOUN	munt/NOUN
kruid/NOUN	leven/NOUN	manifestatie/NOUN	mutts/NOUN
kruidenbitter/NOUN	levenspartner/NOUN	mannetje/NOUN	muur/NOUN
kruim/NOUN	lexicon/NOUN	manoeuvre/NOUN	muziekstuk/NOUN
kruin/NOUN	lezer/NOUN	markt/NOUN	mystificatie/NOUN
kruis/NOUN	lezing/NOUN	Mars/NOUN	naakt/NOUN
kruising/NOUN	liberaal/NOUN	masker/NOUN	naald/NOUN
kruisweg/NOUN	licentie/NOUN	massa/NOUN	nachtwacht/NOUN
kruk/NOUN	lichaam/NOUN	mast/NOUN	nagel/NOUN
krul/NOUN	licht/NOUN	master/NOUN	naspel/NOUN
kuif/NOUN	lichtgewicht/NOUN	mastiek/NOUN	navigatie/NOUN
kuit/NOUN	lichting/NOUN	mat/NOUN	nederigheid/NOUN
kunst/NOUN	lid/NOUN	materie/NOUN	neerslag/NOUN
kurk/NOUN	lied/NOUN	matje/NOUN	nepper/NOUN
kwaad/NOUN	Lier/NOUN	matrix/NOUN	nest/NOUN
kwak/NOUN	lift/NOUN	ME/NOUN	net/NOUN
kwalificatie/NOUN	lijden/NOUN	medium/NOUN	netheid/NOUN
kwart/NOUN	lijfje/NOUN	mee/NOUN	nevel/NOUN
kwartier/NOUN	lijn/NOUN	meester/NOUN	nicht/NOUN
kwast/NOUN	lijst/NOUN	meisje/NOUN	nood/NOUN
kweek/NOUN	lik/NOUN	memorandum/NOUN	noot/NOUN
kweeper/NOUN	linie/NOUN	memorie/NOUN	nummer/NOUN
kwelwater/NOUN	linkerzijde/NOUN	menu/NOUN	nummertje/NOUN
kwestie/NOUN	lintje/NOUN	meter/NOUN	object/NOUN
label/NOUN	lob/NOUN	micrometer/NOUN	offer/NOUN
ladder/NOUN	locatie/NOUN	middel/NOUN	officier/NOUN
lading/NOUN	loge/NOUN	middelpunt/NOUN	omber/NOUN
lak/NOUN	lompen/NOUN	migratie/NOUN	omgang/NOUN
lama/NOUN	lood/NOUN	min/NOUN	omhaal/NOUN
lamlendigheid/NOUN	loop/NOUN	mineraal/NOUN	omloop/NOUN
lamp/NOUN	loper/NOUN	minuut/NOUN	omnibus/NOUN
land/NOUN	lot/NOUN	missie/NOUN	omslag/NOUN
landschap/NOUN	lucht/NOUN	model/NOUN	omtrek/NOUN
last/NOUN	luik/NOUN	module/NOUN	omzetting/NOUN
latex/NOUN	luim/NOUN	modus/NOUN	onafhankelijkheid/NOUN
lector/NOUN	lul/NOUN	moer/NOUN	onderdruk/NOUN
leer/NOUN	lust/NOUN	mol/NOUN	onderhoud/NOUN
leeuw/NOUN	lusteloosheid/NOUN	molen/NOUN	onderkruiper/NOUN



ondervraging/NOUN	overspanning/NOUN	piëteit/NOUN	prefect/NOUN
onderwerp/NOUN	overweging/NOUN	pieterman/NOUN	prelude/NOUN
oneindigheid/NOUN	paal/NOUN	pijl/NOUN	premie/NOUN
ongerechtigheid/NOUN	paar/NOUN	pijp/NOUN	prent/NOUN
onregelmatigheid/NOUN	paard/NOUN	pijpje/NOUN	presentatie/NOUN
ontbinding/NOUN	paardensprong/NOUN	pik/NOUN	prijs/NOUN
onthouding/NOUN	pad/NOUN	piket/NOUN	prik/NOUN
ontlasting/NOUN	paddestoel/NOUN	pil/NOUN	prikkel/NOUN
ontmoeting/NOUN	pak/NOUN	pin/NOUN	primaat/NOUN
ontrouw/NOUN	pakje/NOUN	pink/NOUN	principe/NOUN
ontsteking/NOUN	paladijn/NOUN	piste/NOUN	prins/NOUN
ontvangst/NOUN	Pan/NOUN	pit/NOUN	procedure/NOUN
ontwikkeling/NOUN	pand/NOUN	plaag/NOUN	product/NOUN
ontzag/NOUN	pantser/NOUN	plaat/NOUN	productie/NOUN
ontzetting/NOUN	pap/NOUN	plaatje/NOUN	proef/NOUN
oog/NOUN	papegaai/NOUN	plaats/NOUN	profiel/NOUN
oogst/NOUN	papier/NOUN	plak/NOUN	programma/NOUN
oordeel/NOUN	paprika/NOUN	plakker/NOUN	project/NOUN
oorveeg/NOUN	paradijs/NOUN	plan/NOUN	projectie/NOUN
opbouw/NOUN	parallel/NOUN	plantenspuit/NOUN	proloog/NOUN
opdonder/NOUN	parel/NOUN	plas/NOUN	promotie/NOUN
opdracht/NOUN	parool/NOUN	plateau/NOUN	promotor/NOUN
opening/NOUN	partij/NOUN	plek/NOUN	propositie/NOUN
opera/NOUN	partner/NOUN	ploeg/NOUN	protagonist/NOUN
operatie/NOUN	pas/NOUN	pluim/NOUN	provincie/NOUN
opgave/NOUN	passage/NOUN	pluis/NOUN	provisie/NOUN
opheffing/NOUN	pastel/NOUN	plus/NOUN	provoost/NOUN
opkomst/NOUN	pastorale/NOUN	poeder/NOUN	pruim/NOUN
oplichting/NOUN	paternoster/NOUN	poeier/NOUN	publicatie/NOUN
oplossing/NOUN	patriarch/NOUN	poes/NOUN	publiek/NOUN
opname/NOUN	patroon/NOUN	politiek/NOUN	puinhoop/NOUN
oppositie/NOUN	paviljoen/NOUN	polo/NOUN	pulp/NOUN
opslag/NOUN	peer/NOUN	pols/NOUN	punk/NOUN
opstelling/NOUN	pegel/NOUN	pool/NOUN	punt/NOUN
optie/NOUN	peil/NOUN	poort/NOUN	pupil/NOUN
optreden/NOUN	PEN/NOUN	poot/NOUN	put/NOUN
orakel/NOUN	pens/NOUN	pop/NOUN	raad/NOUN
oranje/NOUN	pension/NOUN	populatie/NOUN	raadsheer/NOUN
orde/NOUN	peper/NOUN	portefeuille/NOUN	raam/NOUN
organisatie/NOUN	performance/NOUN	portret/NOUN	raamwerk/NOUN
Oriënt/NOUN	Pers/NOUN	pose/NOUN	rad/NOUN
ouderdom/NOUN	persoon/NOUN	positie/NOUN	radio/NOUN
oudste/NOUN	perspectief/NOUN	post/NOUN	RAM/NOUN
ouwe/NOUN	pest/NOUN	postpakket/NOUN	rand/NOUN
overdruk/NOUN	piccolo/NOUN	pot/NOUN	rank/NOUN
overgang/NOUN	piek/NOUN	pr/NOUN	ratel/NOUN
overgangsjaren/NOUN	pieper/NOUN	praam/NOUN	ratio/NOUN
overmacht/NOUN	pier/NOUN	praatje/NOUN	rationalisme/NOUN
overslag/NOUN	Piet/NOUN	praktijk/NOUN	rattenkop/NOUN



rayon/NOUN	rijm/NOUN	schets/NOUN	slagpen/NOUN
razernij/NOUN	ring/NOUN	scheur/NOUN	slak/NOUN
reactie/NOUN	roddelpraat/NOUN	scheut/NOUN	slakkenhuis/NOUN
realisme/NOUN	roede/NOUN	schijf/NOUN	Slang/NOUN
rebellie/NOUN	roek/NOUN	schijnsel/NOUN	slaper/NOUN
receptie/NOUN	Rok/NOUN	schild/NOUN	slapte/NOUN
recht/NOUN	rol/NOUN	schim/NOUN	slee/NOUN
rechtervleugel/NOUN	roller/NOUN	schipbreuk/NOUN	sleep/NOUN
rechtschapenheid/NOUN	ronde/NOUN	schoft/NOUN	sleper/NOUN
reclame/NOUN	ronding/NOUN	schok/NOUN	sleutel/NOUN
rector/NOUN	rondje/NOUN	schol/NOUN	slib/NOUN
rede/NOUN	rondrit/NOUN	school/NOUN	slinger/NOUN
reductie/NOUN	rooie/NOUN	schoorsteen/NOUN	slippendrager/NOUN
ree/NOUN	roos/NOUN	schoot/NOUN	slokop/NOUN
reep/NOUN	rot/NOUN	schorpioen/NOUN	sloof/NOUN
regel/NOUN	rotting/NOUN	Schot/NOUN	sloop/NOUN
regeneratie/NOUN	rouw/NOUN	schotel/NOUN	slot/NOUN
regent/NOUN	rozenkrans/NOUN	schouderstuk/NOUN	sluiting/NOUN
regering/NOUN	RUG/NOUN	schouw/NOUN	slurf/NOUN
regime/NOUN	ruggengraat/NOUN	schrift/NOUN	smaak/NOUN
register/NOUN	ruim/NOUN	schrijven/NOUN	smak/NOUN
rehabilitatie/NOUN	ruimte/NOUN	schrik/NOUN	snavel/NOUN
rei/NOUN	ruit/NOUN	schroot/NOUN	snede/NOUN
reikwijdte/NOUN	run/NOUN	schuif/NOUN	snee/NOUN
reiniging/NOUN	runner/NOUN	schuiver/NOUN	sneeuw/NOUN
rekening/NOUN	rups/NOUN	schurk/NOUN	sneeuwbal/NOUN
rekensom/NOUN	rust/NOUN	sectie/NOUN	snoer/NOUN
relatie/NOUN	rustplaats/NOUN	segment/NOUN	snotaap/NOUN
rem/NOUN	sabel/NOUN	selectie/NOUN	snotneus/NOUN
ren/NOUN	saffier/NOUN	serie/NOUN	sok/NOUN
repercussie/NOUN	salon/NOUN	serpent/NOUN	solo/NOUN
repetitie/NOUN	samenstelling/NOUN	service/NOUN	som/NOUN
repetitor/NOUN	satelliet/NOUN	set/NOUN	sommatie/NOUN
repressie/NOUN	scanner/NOUN	sfeer/NOUN	sonde/NOUN
reprise/NOUN	scène/NOUN	shit/NOUN	soort/NOUN
reproductie/NOUN	schaak/NOUN	sifon/NOUN	souvenir/NOUN
requiem/NOUN	schaal/NOUN	sigaret/NOUN	span/NOUN
reserve/NOUN	schaamtegevoel/NOUN	sightseeing/NOUN	spanning/NOUN
residentie/NOUN	schaap/NOUN	signatuur/NOUN	specie/NOUN
resolutie/NOUN	schaar/NOUN	singel/NOUN	speculatie/NOUN
ressort/NOUN	schacht/NOUN	single/NOUN	speen/NOUN
revisie/NOUN	schans/NOUN	sint/NOUN	spel/NOUN
rib/NOUN	schat/NOUN	situatie/NOUN	speldenknop/NOUN
ridder/NOUN	scheepswand/NOUN	sjabloon/NOUN	speler/NOUN
riem/NOUN	scheiding/NOUN	sla/NOUN	spies/NOUN
riet/NOUN	schelp/NOUN	slaaf/NOUN	spin/NOUN
rij/NOUN	schema/NOUN	slaap/NOUN	spinnenkop/NOUN
rijk/NOUN	scherm/NOUN	slag/NOUN	spiraal/NOUN
rijkdom/NOUN	scherpte/NOUN	slager/NOUN	spits/NOUN



split/NOUN	stok/NOUN	televisie/NOUN	trio/NOUN
splitsing/NOUN	stoker/NOUN	telex/NOUN	troep/NOUN
spontaniteit/NOUN	stoot/NOUN	temperatuur/NOUN	trojka/NOUN
spook/NOUN	stop/NOUN	tenor/NOUN	trommel/NOUN
spoor/NOUN	straal/NOUN	tent/NOUN	TROS/NOUN
spot/NOUN	straat/NOUN	term/NOUN	trots/NOUN
spriet/NOUN	straf/NOUN	termijn/NOUN	tuimelaar/NOUN
sprong/NOUN	streek/NOUN	terracotta/NOUN	turf/NOUN
spruit/NOUN	streep/NOUN	test/NOUN	turkoois/NOUN
sput/NOUN	strik/NOUN	testament/NOUN	tussenstation/NOUN
staak/NOUN	strip/NOUN	textiel/NOUN	type/NOUN
staal/NOUN	stroming/NOUN	theater/NOUN	typograaf/NOUN
staart/NOUN	stroom/NOUN	thee/NOUN	uil/NOUN
staat/NOUN	strop/NOUN	theeblad/NOUN	uilenbal/NOUN
stadswijk/NOUN	stropop/NOUN	theewater/NOUN	uitbarsting/NOUN
staf/NOUN	struweel/NOUN	thema/NOUN	uitdrukking/NOUN
stam/NOUN	studie/NOUN	thesis/NOUN	uitgave/NOUN
stamper/NOUN	studio/NOUN	tijd/NOUN	uitgebreidheid/NOUN
stand/NOUN	stuk/NOUN	tik/NOUN	uitgever/NOUN
standaard/NOUN	substantie/NOUN	tip/NOUN	uithaal/NOUN
standplaats/NOUN	suggestie/NOUN	titan/NOUN	uitkomst/NOUN
stap/NOUN	suiker/NOUN	titel/NOUN	uitloop/NOUN
stapel/NOUN	superieur/NOUN	tocht/NOUN	uitloper/NOUN
station/NOUN	supplement/NOUN	toer/NOUN	uitmonstering/NOUN
status/NOUN	systeem/NOUN	toeter/NOUN	uitslag/NOUN
steek/NOUN	taak/NOUN	toets/NOUN	uitsluiting/NOUN
steekspel/NOUN	taart/NOUN	toilet/NOUN	uitsmijter/NOUN
steen/NOUN	tabernakel/NOUN	ton/NOUN	uitspraak/NOUN
steenbok/NOUN	tableau/NOUN	toneel/NOUN	uitvaart/NOUN
steentje/NOUN	tablet/NOUN	toog/NOUN	uitval/NOUN
stek/NOUN	tactiek/NOUN	toon/NOUN	uitvoer/NOUN
stelling/NOUN	tafel/NOUN	top/NOUN	uitvoering/NOUN
stem/NOUN	taille/NOUN	topper/NOUN	uitweg/NOUN
stemming/NOUN	tak/NOUN	totalisator/NOUN	uitwerking/NOUN
stempel/NOUN	taks/NOUN	tour/NOUN	uitzending/NOUN
step/NOUN	talud/NOUN	tracé/NOUN	uitzicht/NOUN
STER/NOUN	talud/NOUN	trailer/NOUN	umlaut/NOUN
sterrenkijker/NOUN	tand/NOUN	transcriptie/NOUN	unit/NOUN
sterretje/NOUN	tang/NOUN	transfer/NOUN	uur/NOUN
steun/NOUN	tank/NOUN	transmissie/NOUN	vaandel/NOUN
stier/NOUN	TAP/NOUN	transport/NOUN	vaart/NOUN
stierennek/NOUN	tas/NOUN	transporteur/NOUN	vacht/NOUN
stift/NOUN	techniek/NOUN	trap/NOUN	vader/NOUN
stijfheid/NOUN	teelt/NOUN	trede/NOUN	vak/NOUN
stijl/NOUN	teen/NOUN	treffen/NOUN	val/NOUN
stip/NOUN	teken/NOUN	trek/NOUN	variatie/NOUN
stoet/NOUN	tekening/NOUN	trekker/NOUN	vastberadenheid/NOUN
stof/NOUN	tekort/NOUN	trens/NOUN	vat/NOUN
stofdoek/NOUN	tel/NOUN	treurigheid/NOUN	vechter/NOUN



veer/NOUN	vlaggenschip/NOUN	waardigheid/NOUN	zee/NOUN
vel/NOUN	vlaggenstok/NOUN	waarneming/NOUN	zee-engte/NOUN
veld/NOUN	vlak/NOUN	wacht/NOUN	zegel/NOUN
velletje/NOUN	vlegel/NOUN	wafel/NOUN	zegen/NOUN
vendu/NOUN	vlek/NOUN	wagen/NOUN	zeil/NOUN
venster/NOUN	vleugel/NOUN	wal/NOUN	zekerheid/NOUN
verbeelding/NOUN	vlies/NOUN	wand/NOUN	zelfbediening/NOUN
verbinding/NOUN	vloed/NOUN	wandaad/NOUN	zender/NOUN
verbranding/NOUN	vloedgolf/NOUN	wandelstok/NOUN	zending/NOUN
verdediger/NOUN	vloer/NOUN	want/NOUN	zenuwcentrum/NOUN
verdediging/NOUN	vloot/NOUN	wapening/NOUN	zet/NOUN
verdedigingslinie/NOUN	vlucht/NOUN	was/NOUN	zetel/NOUN
verdichting/NOUN	vod/NOUN	water/NOUN	zever/NOUN
verduistering/NOUN	voeding/NOUN	WC/NOUN	zicht/NOUN
vergelijking/NOUN	voet/NOUN	Weegschaal/NOUN	ziel/NOUN
vergroting/NOUN	voetbal/NOUN	weer/NOUN	zij/NOUN
verhoging/NOUN	voetstap/NOUN	weerstand/NOUN	zijde/NOUN
verhouding/NOUN	volk/NOUN	weg/NOUN	zijn/NOUN
verkenner/NOUN	vondst/NOUN	wereld/NOUN	zilver/NOUN
verklaring/NOUN	voordeel/NOUN	werk/NOUN	zin/NOUN
verkleining/NOUN	voordracht/NOUN	werking/NOUN	zitting/NOUN
verkoop/NOUN	voorhoede/NOUN	werkloosheid/NOUN	zolder/NOUN
verkoping/NOUN	voorspel/NOUN	werplijn/NOUN	zoldering/NOUN
Verlichting/NOUN	voorstelling/NOUN	westen/NOUN	zomer/NOUN
verlies/NOUN	vooruitgang/NOUN	wet/NOUN	zon/NOUN
verloop/NOUN	vordering/NOUN	weten/NOUN	zondagskind/NOUN
verlossing/NOUN	vork/NOUN	wetsontwerp/NOUN	zone/NOUN
vermogen/NOUN	vorm/NOUN	wezen/NOUN	zool/NOUN
verpaupering/NOUN	vorst/NOUN	wijs/NOUN	zoon/NOUN
verplaatsing/NOUN	vos/NOUN	winst/NOUN	zorg/NOUN
vers/NOUN	vossenjacht/NOUN	wip/NOUN	zot/NOUN
versterf/NOUN	vraag/NOUN	wissel/NOUN	zucht/NOUN
versterking/NOUN	vreemdeling/NOUN	wit/NOUN	zuidpool/NOUN
vertegenwoordiger/NOUN	vrees/NOUN	woord/NOUN	zuiger/NOUN
vertrek/NOUN	vreten/NOUN	worm/NOUN	zuivering/NOUN
vervolging/NOUN	vreugde/NOUN	wortel/NOUN	zuster/NOUN
verzamelnaam/NOUN	vriend/NOUN	wraakneming/NOUN	zuur/NOUN
verzekering/NOUN	vrijbuiters/NOUN	y/NOUN	zwaard/NOUN
verzet/NOUN	vrijstelling/NOUN	zaad/NOUN	zwaarte/NOUN
vest/NOUN	vrolijkheid/NOUN	zaag/NOUN	zwaartepunt/NOUN
vet/NOUN	vrouw/NOUN	zaak/NOUN	zwaluwstaart/NOUN
video/NOUN	vrouwtje/NOUN	zadel/NOUN	zwanenhals/NOUN
vijg/NOUN	vuil/NOUN	zak/NOUN	zweet/NOUN
vijzel/NOUN	vuiligheid/NOUN	zakenrelatie/NOUN	zwendel/NOUN
vinger/NOUN	vuur/NOUN	zand/NOUN	zwieper/NOUN
Vis/NOUN	vuurtoren/NOUN	zang/NOUN	zwier/NOUN
visitatie/NOUN	Waal/NOUN	zanger/NOUN	zwijn/NOUN
visite/NOUN	waard/NOUN	zebra/NOUN	zwik/NOUN
vizier/NOUN	waarde/NOUN	zedes/NOUN	



aangelegd/ADJ	bezig/ADJ	fel/ADJ	kleingeestig/ADJ
aangenaam/ADJ	biologisch/ADJ	ferm/ADJ	kleinschalig/ADJ
aangenomen/ADJ	bits/ADJ	fijn/ADJ	kleinzielig/ADJ
aangepast/ADJ	bitter/ADJ	fit/ADJ	knap/ADJ
aanspreekbaar/ADJ	blauw/ADJ	flink/ADJ	knellig/ADJ
aantrekkelijk/ADJ	bleek/ADJ	fout/ADJ	komisch/ADJ
aanwezig/ADJ	blijmoedig/ADJ	geblokkeerd/ADJ	kort/ADJ
aanzienlijk/ADJ	bloemig/ADJ	gebonden/ADJ	kortademig/ADJ
aardig/ADJ	boos/ADJ	gebroken/ADJ	koud/ADJ
aards/ADJ	boosaardig/ADJ	geciviliseerd/ADJ	kras/ADJ
abstract/ADJ	bot/ADJ	geel/ADJ	kritiek/ADJ
accuraat/ADJ	bourgeois/ADJ	geleerd/ADJ	krukkig/ADJ
actueel/ADJ	breed/ADJ	gemakkelijk/ADJ	kwetsbaar/ADJ
adembenemend/ADJ	buitenlands/ADJ	genoegzaam/ADJ	laag/ADJ
afgemeten/ADJ	calorierijk/ADJ	gepeperd/ADJ	lam/ADJ
afgodisch/ADJ	clean/ADJ	geremd/ADJ	lang/ADJ
afgrijselijk/ADJ	complex/ADJ	gericht/ADJ	langzaam/ADJ
afhankelijk/ADJ	demonisch/ADJ	getapt/ADJ	lastig/ADJ
afschuwwekkend/ADJ	depressief/ADJ	getekend/ADJ	leeg/ADJ
afstandelijk/ADJ	direct/ADJ	getroffen/ADJ	levend/ADJ
afstotelijk/ADJ	dol/ADJ	gevaarlijk/ADJ	licht/ADJ
akkoord/ADJ	dom/ADJ	geweldig/ADJ	lief/ADJ
allergisch/ADJ	donker/ADJ	gewoon/ADJ	lijvig/ADJ
alternatief/ADJ	dood/ADJ	gezond/ADJ	los/ADJ
amateuristisch/ADJ	doorkneed/ADJ	gigantisch/ADJ	luchthartig/ADJ
antipathiek/ADJ	driftig/ADJ	goedkoop/ADJ	luchtig/ADJ
apocalyptisch/ADJ	droevig/ADJ	grandioos/ADJ	lucide/ADJ
arm/ADJ	dronken/ADJ	grappig/ADJ	lui/ADJ
armzalig/ADJ	droog/ADJ	grijs/ADJ	misselijkmakend/ADJ
bang/ADJ	druilerig/ADJ	groen/ADJ	modern/ADJ
bedeesd/ADJ	druk/ADJ	groot/ADJ	moeilijk/ADJ
bedompt/ADJ	duidelijk/ADJ	handig/ADJ	mogelijk/ADJ
bedroevend/ADJ	dun/ADJ	heel/ADJ	naakt/ADJ
bedrukt/ADJ	duur/ADJ	heet/ADJ	naar/ADJ
beduidend/ADJ	edel/ADJ	heilig/ADJ	nauwkeurig/ADJ
behouden/ADJ	eendaags/ADJ	helder/ADJ	neerslachtig/ADJ
bejaard/ADJ	eenjarig/ADJ	hoffelijk/ADJ	negatief/ADJ
belangrijk/ADJ	eenstemmig/ADJ	hol/ADJ	nieuw/ADJ
beleggen/ADJ	eenvoudig/ADJ	hoog/ADJ	nieuwerwets/ADJ
benieuwd/ADJ	eerlijk/ADJ	illegaal/ADJ	nijpend/ADJ
bereden/ADJ	effen/ADJ	impulsief/ADJ	normaal/ADJ
berekend/ADJ	eigen/ADJ	intelligent/ADJ	nuchter/ADJ
beschaafd/ADJ	eigentijds/ADJ	jong/ADJ	omvangrijk/ADJ
beschikbaar/ADJ	elektrisch/ADJ	juist/ADJ	onaangenaam/ADJ
beschroomd/ADJ	enorm/ADJ	keihard/ADJ	onaardig/ADJ
bestand/ADJ	enthousiast/ADJ	keurig/ADJ	onbeheerst/ADJ
betrokken/ADJ	ernstig/ADJ	klef/ADJ	onbeslist/ADJ
beurs/ADJ	ervaren/ADJ	klein/ADJ	ongebonden/ADJ
	fatsoenlijk/ADJ	kleinburgerlijk/ADJ	ongekunsteld/ADJ



ongezond/ADJ	snel/ADJ	vroeger/ADJ	afspelen/VERB
oninteressant/ADJ	snoezig/ADJ	vrolijk/ADJ	afvragen/VERB
onterecht/ADJ	sociaal/ADJ	vruchtbaar/ADJ	afwijken/VERB
onthand/ADJ	spitsvondig/ADJ	waard/ADJ	antwoorden/VERB
onweerlegbaar/ADJ	sportief/ADJ	walgelig/ADJ	attenderen/VERB
onzacht/ADJ	sprekend/ADJ	warm/ADJ	barsten/VERB
onzuiver/ADJ	stads/ADJ	weezinwekkend/ADJ	bedanken/VERB
opgewekt/ADJ	sterk/ADJ	wijd/ADJ	bedekken/VERB
opvallend/ADJ	stijf/ADJ	wijs/ADJ	bedenken/VERB
oranje/ADJ	stout/ADJ	wit/ADJ	bedoelen/VERB
oud/ADJ	straight/ADJ	zat/ADJ	bedragen/VERB
paars/ADJ	strategisch/ADJ	zeer/ADJ	bedreigen/VERB
paradijselijk/ADJ	stuitend/ADJ	zilt/ADJ	bedrijven/VERB
passend/ADJ	superieur/ADJ	zilveren/ADJ	begeleiden/VERB
persistent/ADJ	sympathiek/ADJ	zout/ADJ	beginnen/VERB
persoonlijk/ADJ	taai/ADJ	zuur/ADJ	begraven/VERB
pijnlijk/ADJ	tactisch/ADJ	zwaar/ADJ	begrijpen/VERB
plezant/ADJ	tam/ADJ	zwaarmoedig/ADJ	begroeten/VERB
politiek/ADJ	teruggetrokken/ADJ	zwart/ADJ	behoeven/VERB
populair/ADJ	toornig/ADJ		behouden/VERB
positief/ADJ	traag/ADJ	aanbrengen/VERB	beïnvloeden/VERB
prachtig/ADJ	treurig/ADJ	aandoen/VERB	bekennen/VERB
prettig/ADJ	triest/ADJ	aanduiden/VERB	bekijken/VERB
prima/ADJ	tweeslachtig/ADJ	aangaan/VERB	beleven/VERB
primair/ADJ	uitgeput/ADJ	aangeven/VERB	bellen/VERB
publiek/ADJ	uitstekend/ADJ	aanhouden/VERB	beloven/VERB
puik/ADJ	vaag/ADJ	aankijken/VERB	bemoeien/VERB
rank/ADJ	vals/ADJ	aankomen/VERB	benaderen/VERB
realistisch/ADJ	ver/ADJ	aannemen/VERB	benoemen/VERB
rijk/ADJ	verlegen/ADJ	aanraken/VERB	beoordelen/VERB
rins/ADJ	verlicht/ADJ	aansluiten/VERB	bepalen/VERB
rond/ADJ	verloren/ADJ	aantasten/VERB	beperken/VERB
rood/ADJ	vermeldenswaard/ADJ	aantonen/VERB	bereiken/VERB
ruig/ADJ	vers/ADJ	aantreffen/VERB	beroemen/VERB
rustig/ADJ	verstandig/ADJ	aantrekken/VERB	berusten/VERB
ruw/ADJ	vertoord/ADJ	aanvaarden/VERB	beschermen/VERB
saai/ADJ	vervelend/ADJ	aanwijzen/VERB	beschikken/VERB
scabreus/ADJ	vet/ADJ	aanzien/VERB	beschrijven/VERB
schaapachtig/ADJ	vierkant/ADJ	aarzelen/VERB	beseffen/VERB
scherp/ADJ	vies/ADJ	accepteren/VERB	beslissen/VERB
schoon/ADJ	vliegend/ADJ	achten/VERB	besluiten/VERB
schraal/ADJ	vlug/ADJ	achterlaten/VERB	bespreken/VERB
schrande/ADJ	vol/ADJ	afhalen/VERB	besteden/VERB
simpel/ADJ	volgend/ADJ	afleggen/VERB	bestellen/VERB
slaapverwekkend/ADJ	volumineus/ADJ	afleiden/VERB	bestemmen/VERB
slap/ADJ	vreemd/ADJ	aflopen/VERB	bestrijden/VERB
smakeloos/ADJ	vriendelijk/ADJ	afnemen/VERB	bestuderen/VERB
smal/ADJ	vrij/ADJ	afpassen/VERB	betalen/VERB
smerig/ADJ	vroeg/ADJ	afsluiten/VERB	betreffen/VERB



betrekken/VERB	doorbrengen/VERB	grinniken/VERB	komen/VERB
bevalen/VERB	doordringen/VERB	groeien/VERB	kopen/VERB
bevatten/VERB	doorgaan/VERB	groeten/VERB	kosten/VERB
bevelen/VERB	draaien/VERB	grommen/VERB	kraken/VERB
beven/VERB	dragen/VERB	gunnen/VERB	kreunen/VERB
bevestigen/VERB	drijven/VERB	halen/VERB	krijgen/VERB
bevinden/VERB	dringen/VERB	handelen/VERB	kruijen/VERB
bevorderen/VERB	drinken/VERB	handhaven/VERB	kunnen/VERB
bevredigen/VERB	drogen/VERB	hangen/VERB	kussen/VERB
bevrijden/VERB	dromen/VERB	hanteren/VERB	kweken/VERB
bewaren/VERB	drukken/VERB	haten/VERB	lachen/VERB
bewegen/VERB	duiden/VERB	hebben/VERB	laten/VERB
beweren/VERB	duiken/VERB	hechten/VERB	leggen/VERB
bewijzen/VERB	duimen/VERB	heffen/VERB	leiden/VERB
bewonderen/VERB	duizelen/VERB	helpen/VERB	lenen/VERB
bezighouden/VERB	duren/VERB	herhalen/VERB	leren/VERB
bezitten/VERB	durven/VERB	herstellen/VERB	letten/VERB
bezoeken/VERB	dutten/VERB	heten/VERB	leunen/VERB
bezorgen/VERB	duwen/VERB	hijgen/VERB	leven/VERB
bidden/VERB	dwingen/VERB	hoeven/VERB	leveren/VERB
bieden/VERB	eindigen/VERB	hollen/VERB	lezen/VERB
binden/VERB	eisen/VERB	hopen/VERB	lichten/VERB
binnenkomen/VERB	ergeren/VERB	horen/VERB	liegen/VERB
blazen/VERB	erkennen/VERB	houden/VERB	liggen/VERB
blijken/VERB	ervaren/VERB	huilen/VERB	lijden/VERB
blijven/VERB	eten/VERB	informereren/VERB	lopen/VERB
bloeien/VERB	falen/VERB	ingaan/VERB	loslaten/VERB
boeien/VERB	fluisteren/VERB	ingrijpen/VERB	luiden/VERB
bouwen/VERB	fluiten/VERB	inhouden/VERB	luisteren/VERB
branden/VERB	formuleren/VERB	inladen/VERB	lukken/VERB
breken/VERB	functioneren/VERB	innemen/VERB	maken/VERB
brengen/VERB	gaan/VERB	inrichten/VERB	meebrengen/VERB
brullen/VERB	gebeuren/VERB	instellen/VERB	meemaken/VERB
buigen/VERB	gebruiken/VERB	interessereren/VERB	meenemen/VERB
bukken/VERB	gedragen/VERB	jagen/VERB	melden/VERB
citeren/VERB	gelden/VERB	kenmerken/VERB	menen/VERB
concentreren/VERB	geloven/VERB	kennen/VERB	mengen/VERB
concluderen/VERB	genezen/VERB	keren/VERB	merken/VERB
confronteren/VERB	genieten/VERB	kiezen/VERB	meten/VERB
controleren/VERB	geschieden/VERB	kijken/VERB	mislukken/VERB
dalen/VERB	getuigen/VERB	klagen/VERB	missen/VERB
dansen/VERB	geven/VERB	kleden/VERB	moeten/VERB
deelnemen/VERB	gillen/VERB	klimmen/VERB	mogen/VERB
dekken/VERB	glanzen/VERB	klinken/VERB	mompelen/VERB
delen/VERB	glijden/VERB	kloppen/VERB	nadenken/VERB
denken/VERB	glimlachen/VERB	knikken/VERB	naderen/VERB
dienen/VERB	gooien/VERB	knippen/VERB	nagaan/VERB
doceren/VERB	grijzen/VERB	koesteren/VERB	nalaten/VERB
doen/VERB	grijpen/VERB	koken/VERB	neerleggen/VERB



nemen/VERB	redden/VERB	stukgaan/VERB	vertrekken/VERB
noemen/VERB	regelen/VERB	tegenkomen/VERB	verzamelen/VERB
noteren/VERB	reizen/VERB	tekenen/VERB	verzinnen/VERB
oefenen/VERB	rekenen/VERB	tellen/VERB	vinden/VERB
omdraaien/VERB	richten/VERB	terugkomen/VERB	vliegen/VERB
omgaan/VERB	rijden/VERB	tieren/VERB	voelen/VERB
omkeren/VERB	roepen/VERB	tonen/VERB	voeren/VERB
omschrijven/VERB	roken/VERB	treffen/VERB	volgen/VERB
omvatten/VERB	schelen/VERB	trekken/VERB	voorkomen/VERB
onderbreken/VERB	schieten/VERB	trouwen/VERB	voorstellen/VERB
onderzoeken/VERB	schilderen/VERB	uitblazen/VERB	voorzien/VERB
onthouden/VERB	schrijven/VERB	uitgaan/VERB	vormen/VERB
ontkoppelen/VERB	schuiven/VERB	uitkomen/VERB	vragen/VERB
ontmoeten/VERB	slaan/VERB	uitleggen/VERB	vrezen/VERB
ontvangen/VERB	slagen/VERB	uitslijken/VERB	vullen/VERB
ontwikkelen/VERB	slapen/VERB	uitvoeren/VERB	wachten/VERB
openen/VERB	sluiten/VERB	uitzaaien/VERB	wagen/VERB
opereren/VERB	smeren/VERB	vallen/VERB	wassen/VERB
oplossen/VERB	snappen/VERB	vangen/VERB	wennen/VERB
opnemen/VERB	spelen/VERB	veranderen/VERB	werken/VERB
optreden/VERB	spreken/VERB	verbinden/VERB	wijzen/VERB
opvatten/VERB	springen/VERB	verbreken/VERB	winnen/VERB
organiseren/VERB	staan/VERB	verdelen/VERB	zeggen/VERB
overnemen/VERB	stangen/VERB	verdienen/VERB	zetten/VERB
pakken/VERB	stappen/VERB	verklaren/VERB	zien/VERB
passen/VERB	steken/VERB	verkopen/VERB	zijn/VERB
plaatsen/VERB	stellen/VERB	verkreukelen/VERB	zingen/VERB
praten/VERB	stemmen/VERB	verlaten/VERB	zitten/VERB
raken/VERB	sterven/VERB	verliezen/VERB	zoeken/VERB
ratelen/VERB	stijgen/VERB	verplichten/VERB	zullen/VERB
reageren/VERB	stoppen/VERB	verschillen/VERB	
realiseren/VERB	studeren/VERB	versmallen/VERB	



7.3 Nouns with 8 or more equivalence relations that are post-edited manually for equivalence, SUMO label and domain label

startstreep:1	spriet:4	meisje:3	zedeloosheid:1	tuidraad:1	willoosheid:1
kaffer:4	weekheid:1	mensenras:1	zoetje:3	uitbundigheid:1	woede:2
vrolijkheid:2	afvaardiging:2	onderhoud:4	zwierbol:1	uitdamping:1	zwak:3
topper:5	beheer:1	toemaatje:2	zwijnjak:1	verbijstering:1	aanrechtblad:1
kwakkelaar:1	couvert:1	wc:1	aanvuring:1	vierde:3	achterstuk:1
acht:1	draadje:1	zaag:1	academie:2	zeekant:3	adjutant:1
tik:3	etage:2	aula:1	afvalhoop:1	zelfkant:3	afstotingskracht:1
blaag:1	knabbeltje:1	inlichting:2	agressiviteit:1	zonwering:1	bast:1
groei:3	kraan:5	lintje:3	automaat:3	zwelgp partij:1	bouwkunst:1
zorgeloosheid:2	magazijnvoorraad:1	ordeketen:1	blooper:1	droogte:1	delging:1
rebellie:1	ruitje:2	rank:1	bloot:3	engel:1	desoriëntatie:1
eindje:2	vreugde:1	rasterwerk:2	bovenzijde:1	gestrengheid:1	discrepantie:1
los:4	waarachtigheid:1	schaamtegevoel:1	directie:1	gom:3	drop:1
uitgave:4	wanprodukt:1	schepje:2	erkenenis:1	grondbeginselen:1	duimzuiger:1
ruit:5	woonstede:1	snoeverij:1	flard:1	gunstbewijs:1	engel:2
zot:2	aanvoer:3	snotaap:1	fraseologie:1	handelmaatschappij:1	fonkeling:1
druif:1	hoofdzetel:1	straat:1	gemoedelijkheid:1	handels huis:1	fraicheur:1
treurigheid:4	hoogachting:1	stroopsmeerder:1	gigant:1	handleiding:1	gezaghebber:1
vastberadenheid:1	ontzag:1	struweel:1	goedaardigheid:1	kaart(e):1	grootwinkelbedrijf:1
verpaupering:1	oorveeg:1	taluud:2	hoofdeinde:2	kladpapiertje:1	houtmijt:1
cadans:1	overvaart:1	tweede:3	kalebas:4	korenwan:1	kattengejank:3
opvulsel:1	piëteit:1	wapenschouwing:1	klokken:1	las:2	klier:1
oplichting:1	reikwijdte:2	wreveligheid:1	kneep:4	maaksel:2	knurf:1
tijdsverloop:1	schraapzucht:1	zwendel:1	kraak:4	nattigheid:2	koppelriem:1
versterking:4	vacht:2	aanzetbuis:1	kroon:2	onbezonnenheid:1	leermeester:4
windmaker:1	vrees:3	attentie:2	lusteloosheid:1	onmenselijkheid:1	lepel:3
barenswee:1	wraakneming:1	dump:1	ontzetting:4	opsteker:1	lichtkrans:1
haakje:1	crisistijd:1	grootheid:2	ootmoed:1	reliëfwerk:1	lieftalligheid:1
schijnsel:2	faciliteit:2	hogeschool:1	overgangsjaren:1	ren:4	lijs:1
spitsheid:1	geestdrijver:1	lied:2	pathos:1	resignatie:1	lijsterbes:1
bijklank:1	kippetje:1	pap:1	pikker:1	rusteloosheid:1	loyaliteit:1
rolletje:1	klepper:1	pluis:2	prospect:1	rustplaats:1	maag:2
vlam:4	letterkast:1	rechtschapenheid:1	pudeur:1	stikkie:1	mandataris:1
eenvoud:3	maniak:1	snebbe:3	reclameplaat:2	structurering:1	onverschrokkenheid:1
fikkie:1	onschuld:5	spitsheid:2	reisroute:1	studentenhuis:1	paprika:3
gewelddadigheid:3	oudste:2	sukkel:1	roeststok:1	trimester:1	portaal:1
gladheid:2	protagonist:1	tegenvoeter:3	sacherijn:1	tumult:1	praatjes:1
negotie:3	roddelpraat:1	toevalligheid:3	schepje:1	uitgeversmaatschappij:1	provincialisme:1
spotje:1	spontaniteit:1	transmutatie:1	schimmel:2	uitspanning:2	ruwheid:2
zwartgalligheid:1	straatloper:1	trekvermogen:1	schuldbrief:1	verbittering:1	schapestal:2
cohort:1	uitweg:1	veerkracht:2	schuldigheid:1	verstandhouding:1	schemertijd:1
godsdienstplechtigheid:1	vechter:1	vermetelheid:1	sprietje:2	vertedering:3	schoolvos:1
kongsi:2	dagdeel:1	vuren hout:2	sprotje:1	vijandelijkheid:1	slokop:1
koppel:5	dikte:4	wijndroesem:1	sterspot:1	voorportaal:1	snotjongen:1
lapje:1	employée	wilskracht:1	stilte:5	wensdroom:1	speldje:3
schurk:2	knoeier:3	woelgeest:1	stuitje:1	werkruimte:1	spotvogel:1



sproei:1	eigendom:2	reserve:5	broosheid:1	opstandigheid:1	stoof:3
studeervtrek:1	flikkering:1	robbenhuid:1	bups:1	opvoeding:1	stopplaats:1
telepaat:1	graveerstift:1	schijn:2	cavalerist:1	opziener:1	stormwind:1
trope:1	haam:3	schoffering:1	centennium:1	orgel:2	storting:1
tube:3	hal:3	sound-track:2	closet:1	paleis:2	strafblad:2
utopist:1	hardvochtigheid:1	spulletjes:1	credo:1	pannenkoekmes:1	strobloem:1
verstuiving:2	havenhoofd:1	stoelleuning:1	drift:1	papegaai:2	strooplikkerij:1
vliegtuigmoederschip:1	hectogram:1	stoorzender:2	eend:2	paprika:2	stropop:4
vredigheid:1	hoekworp:1	stootkussen:2	entrepreneur:1	pekelnat:1	subtiliteit:3
vriendschappelijkheid:1	hypostase:2	stopping:1	gaping:1	perkoenpaal:1	theater:1
wereldburger:1	kattenkop:2	strik:4	gedreun:1	peso:1	toorts:1
zenuwcentrum:1	kelder:1	strooier:1	gerinkel:1	pil:1	tranche:1
zielerust:3	kippenkontje:1	toverkol:1	geschetter:1	pinda:2	uitgaaf:2
zombie:2	klok-en-hamerspel:1	trawant:1	gevoeligheid:3	pindanoot:1	uithangteken:1
zwartekunst:2	knaller:1	tuttebel:1	grafgewelf:1	pol:1	valies:1
aanvechting:1	kopspijker:2	uitdrukking:3	halletje:1	prediker:1	veinzerij:1
accumulator:1	kuip:1	uitgewekene:1	imago:1	prijsberekening:1	verworpen:1
arena:1	kussentje:2	vacht:1	implicatie:1	propriëteit:1	vetheid:2
badinrichting:2	kussentje:3	variabiliteit:1	indispositie:1	rechtlijnigheid:1	vexatie:1
barsheid:1	kwartaal:1	verloftijd:1	inleiding:4	reiswijzer:1	vin:3
bazaar:1	legerplaats:1	vestibule:1	kiesheid:1	reserve:4	viscositeit:1
beeldopbouw:1	lobby:2	vitter:1	kik:1	review:1	vod:4
belangenorganisatie:1	lokaaltje:2	vlagschrift:1	kleed:2	revue:3	volkenbond:1
binnenhof:2	luim:3	voorman:4	klooster:2	rijbaan:1	volmacht:1
boevenbende:1	mandarijntje:1	waard:5	komedie:4	schakeltafel:1	vrouwte:2
bond:2	mastodont:1	waterdam:1	kruizemuntkruid:1	schakering:2	vuilpoes:1
bondgenoot:2	meewarigheid:1	weelde:1	landweg:1	schelp:3	waffel:1
bouwsteen:1	menigte:2	weerschijsel:1	lasverbinding:1	schijf:6	watrgodin:1
buizensysteem:1	ontvangcedel:1	wervelwind:1	leepheid:1	schilletje:2	wederopbloei:1
canapé:1	onzelfzuchtigheid:1	zeereis:1	mantel:2	sensitiviteit:1	weezin:1
cent:1	opgeruimdheid:1	aanstellingskeuring:1	middenstuk:1	shocktoestand:1	weifeling:1
conceptwet:1	pathos:2	achterhuis:1	nieuwkomer:1	slachtoffer:3	werkzaamheid:4
confiscatie:1	peul:3	afbetaling:1	onderscheidingsdrang:1	slib:4	woord:3
coryfee:1	plaveisel:1	afslag:2	ontnuchtering:3	sloompie:1	wrijfpaal:2
dal:1	prentkunst:1	beddenlaken:1	onvriendelijkheid:1	snuisterijen:1	zaalwachter:1
dekblad:1	proces:3	bereidheid:1	onwaarachtigheid:1	spatscher:1	zinspreuk:1
detentie:1	pronker:1	berisping:1	oplettendheid:1	speldenprik:1	zwetskous:1
dinar:1	ratelaar:2	bloedspiegel:1	oproep:1	stemrecht:1	
duikvlucht:1	rayon:4	bocht:3	opsluiting:1	stift:3	



7.4 Verbs with 4 or more equivalence relations that are post-edited manually for equivalence, SUMO label and domain label

doorgaan:5	omwoelen:1	buitenlaten:1	versmallen:3	spoeden:2	afglijden:5
aanhouden:9	opkomen:9	doorreizen:3	vertolken:3	strubbelen:1	afjagen:1
hardlopen:1	opstappen:4	editen:1	verwringen:1	twisten:3	afnemen:8
inkrijgen:1	overlopen:5	manifesteren:2	zinnen:1	uitstromen:2	afroffelen:1
opzetten:15	sjezen:3	opsieren:3	zwaaien:5	uitwijken:4	afsoppen:1
afzetten:21	verbeelden:3	overwinnen:6	zweren:4	verdikken:4	afstrijken:4
afschuiven:7	verminken:4	rondgeven:1	zweren:6	verschonen:4	afstuiten:1
stellen:7	versjacheren:1	tringelen:1	aangrijpen:3	weerklinken:1	afwenden:2
stokken:1	wurmen:2	verblijden:3	afeten:1	wegschieten:2	afzenden:1
expanderen:1	bepakken:1	vlakken:1	afsplijten:1	zwijsen:3	balen:1
tanen:2	bijpassen:1	wiegen:2	afspringen:8	afhollen:2	beroven:2
verleggen:2	omkomen:2	zagen:3	danken:1	afhollen:3	bespeuren:1
stuksnijten:1	opvolgen:4	aankunnen:3	danken:4	afsnellen:1	bezoedelen:2
verzieken:3	overvloeien:3	aansturen:1	doorschuiven:3	afsnellen:3	bijgielen:1
uitdoen:6	ruimen:4	aanzuiveren:1	hijgen:1	afspiegelen:2	blesseren:2
verordineren:1	spinnen:1	afroepen:6	injagen:1	afstrippen:1	boekstaven:1
voorbijlopen:1	visualiseren:1	afsturen:2	invreten:2	bekwamen:2	deprimeren:1
afleggen:4	afspatten:1	converteren:1	inweken:2	beproeven:3	doorstrepen:1
inwerpen:3	afstralen:5	dubben:2	kromlopen:1	disciplineren:1	doorzagen:2
ingaan:4	bekonkelen:1	houden:9	kromlopen:2	doorvlechten:1	dopen:6
uitcijferen:1	doorschuiven:4	invaren:1	lijmen:5	houden:14	grienen:1
unduleren:1	neppen:1	inzenden:2	ontcijferen:2	indraaien:2	ineenzakken:2
voeren:7	opdoffen:1	na-afen:2	opspringen:2	inwalsen:1	inmengen:1
weglopen:6	stuwen:2	omschieten:1	snoeven:1	knoeien:4	klakken:1
zagen:4	toezien:2	onderkennen:1	spoken:5	noteren:3	kleuren:6
afkrijgen:2	uitlaten:4	ontkoppelen:3	uiteten:2	omzetten:5	kokken:1
ergeren:1	uitspringen:4	opdraven:4	uitzuigen:3	ontladen:5	krijten:3
stoken:6	vaarwelzeggen:4	oprollen:5	vervlakken:4	opvreten:2	lichten:7
uitrijden:4	verkwikken:1	opscharrelen:2	vloten:1	overvloeien:5	lijmen:2
volgen:10	verschikken:1	varen:3	waren:2	posten:3	molesteren:1
hutsen:1	waarmaken:1	wegwissen:1	zoemen:1	respecteren:3	omrollen:4
ruggesteunen:1	zeilen:3	afrollen:5	bespioneren:1	rondsnoeven:2	omvaren:3
verwekken:3	afrennen:3	afzadelen:1	doorvoeren:3	rondzenden:1	ontbloten:1
wegwerpen:1	afroepen:5	bijstorten:1	elektriseren:2	schuilgaan:2	ontsluiten:3
achternagaan:2	galmen:1	doorvaren:5	galmen:3	uitroeien:3	opflikkeren:3
afkloppen:3	grabbelen:2	inhaken:3	intreden:3	uitvliegen:2	opklauteren:1
heersen:2	hengsten:4	inhouden:2	kandelaren:1	vallen:10	overwaaen:3
hopen:2	inzetten:6	inlaten:2	kiepen:3	verfoeien:1	poekelen:1
neuzelen:2	meetellen:3	insturen:3	kladderen:1	verkoelen:1	snibben:1
strippen:3	miniseren:1	pagaaien:1	langsgaan:3	verwonderen:4	stabiliseren:1
verpauperen:1	oproepen:6	scholen:1	langslopen:2	viseren:4	stabiliseren:2
volgen:7	prutsen:3	truqueren:1	langstrekken:1	weerklinken:2	stilzitten:3
afscheppen:1	uitspatten:1	uitgieren:1	minnen:1	weiden:2	stofferen:1
doorvorsen:1	afstralen:1	uitwasemen:1	misschieten:1	aansjorren:1	toewerpen:1
inhameren:1	afzwemmen:2	uitweiden:2	nadragen:1	achterhalen:5	uitbannen:1
inhameren:2	afzwemmen:3	uitzeilen:3	omfloersen:1	afdraven:4	uitdrinken:1
omrollen:1	boodschappen:1	vernikkelen:1	renvoyeren:1	afglijden:4	uitkeren:1



venten:2	opsnuiven:1	figureren:4	toedragen:1	doen:8	opsteken:9
verhaasten:1	opwerpen:3	flikken:1	toegeven:7	doordenken:2	optrommelen:1
verkreuteren:1	overrijden:3	formuleren:1	toesnijden:1	doordrijven:2	overkoken:2
verkreuken:1	preluderen:1	geleren:1	traceren:1	doordringen:1	overschrijven:6
vertillen:3	repeteren:3	geselen:2	uiteten:1	doorsnellen:2	overstromen:1
verwonden:2	restitueren:1	gillen:1	uitroeien:2	doorzoeken:3	passeren:6
voorzitten:2	revolteren:2	gillen:3	verdelgen:1	fineren:1	petrificeren:1
vossen:1	rieken:4	harddraven:1	verheerlijken:1	gebeuren:3	platwalsen:2
waggelen:1	romantiseren:1	heiligen:2	verrukken:1	gieren:5	polijsten:2
wriemelen:2	seksen:4	horen:6	vervreemden:3	globaliseren:1	proeven:3
wriemelen:4	shockeren:2	indikken:2	verwaardigen:1	grazen:2	prutsen:1
aanschoppen:1	sissen:1	inkwakken:2	verzinnen:2	halen:3	rekken:4
achterstaan:3	smeden:3	inspringen:4	volproppen:1	harmoniseren:1	renderen:1
afdraven:3	smoezelen:2	instoppen:3	vreten:4	haspelen:2	reorganiseren:1
afhollen:1	sympathiseren:1	jengelen:2	vrijkopen:1	herbergen:3	ringeloren:1
afspringen:10	terugwinnen:2	karteren:1	waarmaken:3	hergroeperen:1	riskeren:2
afspringen:7	terugzien:1	kladderen:2	wegen:3	huppelen:1	robbedoezen:1
afsteken:3	toedenken:3	kleden:4	wegrijden:1	identificeren:6	roezen:1
afstruinen:1	tremmen:1	kleppen:2	aandikken:2	inblazen:1	rollen:14
amortiseren:1	trippen:1	klunen:1	aaneenschakelen:1	ineenschrompelen:1	ronken:4
bedruppen:1	turen:1	kneden:2	aanleren:2	inprenten:2	salariëren:1
bekliederen:1	uitbetalen:2	kwellen:3	aanzuigen:1	instuderen:1	salariëren:2
believen:4	uitkijken:4	lallen:1	afklimmen:1	inwilligen:1	samensmelten:4
bemoeien:1	uitrafelen:2	lijnen:1	afkopen:1	kankeren:1	schaduwen:1
bijgeven:1	vegeteren:2	lobben:1	aflopen:4	klossen:1	schetsen:2
bijleveren:1	verschieten:6	malen:6	afstompen:5	kroelen:1	schreeuwen:4
bomen:1	vervalen:1	neergaan:2	afstoten:1	kwelen:2	schuren:5
danken:2	vetweiden:1	ontrafelen:1	afstrijken:4	kwiteren:1	scoren:5
doemen:1	voeden:7	openlopen:1	afwikkelen:4	lappen:6	scoren:7
doorscheuren:2	vonnissen:1	openrollen:2	agiteren:2	legitimeren:3	scrambelen:1
doorseinen:1	voorlichten:1	opgroeien:1	balanceren:1	loskomen:4	screenen:1
doorvliegen:2	warmlopen:5	overleven:4	bannen:1	meppen:1	settelen:1
doorvliegen:3	zouten:2	overschepen:1	bedrijven:1	minnekozen:1	slaan:8
druppen:2	aandweilen:1	paaien:2	bedrukken:1	misdrijven:1	slempen:1
duikelen:3	achteraangaan:1	platleggen:3	bekomen:2	moffelen:1	slijten:6
exploreren:1	afkletsen:1	plenzen:1	beproeven:2	notificeren:2	smaken:3
fuiven:3	afpellen:1	roeien:4	besnaren:1	ombuigen:3	spieden:1
hullen:2	afrollen:3	rouwen:3	betrekken:5	omdonderen:1	spritsen:1
ingaan:2	afsjouwen:1	rugsteunen:2	bevolken:3	omduikelen:1	starogen:1
joelen:1	afsjouwen:3	sakkeren:1	bevuilen:1	omflikkeren:1	stimuleren:2
kleuren:5	besnoeien:2	schaften:1	bezegelen:1	ommuren:1	stomen:4
kuipen:1	beteren:2	schandvlekken:1	bijschaven:1	omroeren:1	tanden:1
kwadreren:1	betogen:2	schaterlachen:1	bijstorten:2	omvatten:3	tanken:1
louteren:1	bijbenen:1	schreeuwen:5	blazen:2	omwroeten:1	tegenwerpen:1
luiken:2	bijtellen:1	slobberen:3	crediteren:1	omzien:4	teloorgaan:1
meestemmen:1	centrifugereren:1	smakken:3	creperen:3	ontschepen:2	tergen:1
neerstoten:1	coachen:1	snipperen:1	debarkeren:3	ontwrichten:3	terugdeinzen:1
onderbieden:1	dirigeren:3	soppen:1	deduceren:1	opbreken:5	terugwinnen:3
ontgroenen:1	doodlopen:4	speuren:1	dichtklappen:2	opkruien:1	tiranniseren:1
ontwennen:1	doortasten:1	straatslijpen:1	dienen:6	opmarcheren:1	toasten:2
ontzetten:4	doorwaaien:1	terugduwen:1	distilleren:4	opsouperen:1	toelopen:3



toelopen:4	vastpraten:1	verkrampen:1	verschepen:1	voorbestemmen:1	wegdraaien:1
turven:2	verdonkeremanen:1	verluchten:1	versmallen:2	voorbijtrekken:1	wegjagen:1
uitbreken:6	vereelten:1	verpersoonlijken:1	vinden:4	voorleiden:1	wegkopen:1
uitfluiten:2	vereren:1	verplaatsen:4	vlaggen:3	voorspelen:1	wiebelen:2
uitspreken:6	verkorrelen:1	verplanten:1	volbrengen:1	weeklagen:1	zwelgen:2