

# **Comparison of the Final Wordnets German, French, Czech and Estonian**

Version 2, Final  
September 26, 1999

Contributors:

Piek Vossen, University of Amsterdam  
Gerard Escudero, Universitat Politecnica de Catalunya



**Deliverable 2D011D01, WP3, WP4  
EuroWordNet, LE4-8328**

Identification number	LE-4003-2D011D012
Type	Document and Lingware
Title	Comparison of the Final Wordnets Dutch, Spanish and Italian
Status	Draft
Deliverable	2D-011, 2D-012
Work Package	WP3 and WP4
Task	T3.4, T3.5, T4.4, T4.5.
Period covered	April 1999 - July 1999
Date	September 26, 1999
Version	2
Number of pages	33
Authors	⇒ Piek Vossen, University of Amsterdam ⇒ Gerard Escudero, Universitat Politecnica de Catalunya
WP/Task responsible	AVI
Project contact point	Piek Vossen University of Amsterdam Spuistraat 134 1012 VB Amsterdam The Netherlands tel. +31 20 525 4669 fax. +31 20 525 4429 e-mail: <a href="mailto:Piek.Vossen@hum.uva.nl">Piek.Vossen@hum.uva.nl</a>
EC project officer	Ray Hudson
Status	Public
Actual distribution	Project Consortium, the EuroWordNet User Group, the world via <a href="http://www.hum.uva.nl/~ewn">http://www.hum.uva.nl/~ewn</a> .
Supplementary notes	n.a.
Key words	Linguistic Resources, Multilingual Wordnets, Language Engineering

Abstract	This deliverable describes the comparison of the final German, French, Czech and Estonian wordnets. Comparison is done on the basis of the ILI-records that are associated with the synsets.
Status of the abstract	Final
Received on	
Recipient's catalogue number	

## Executive Summary

This deliverable describes the comparison of the final wordnets for German, French, Estonian and Czech. The wordnets contain between 7,5K-22K synsets and 13K-32K word meanings. This is between 8-25% the size of WordNet1.5.

The comparison of the wordnets is done on the basis of the ILI-records to which the synsets refer. There is not a one-to-one mapping of synsets to ILI-records and therefore the comparison is only a rough approximation of the compatibility.

Three types of comparisons have been done:

1. intersection of the associated ILI-records: this indicates the possible translatability of concepts across the languages.
2. the clustering of the associated ILI-records over the EuroWordNet top-ontology: this gives an indication of the conceptual coverage and balancing of the wordnets.
3. the compatibility of hyponymy relations in the wordnets, projected on the associated ILI-records: this gives a rough indication of the similarity in classification structure across the wordnets.

The overall statistics is useful for users of the database to get an idea of the cross-lingual coverage and matching of the data.

## Table of Contents

List of Tables .....	5
1. Introduction .....	7
2. Intersection of the associated ILI-records .....	7
3. The distribution of the associated ILI-records over the top-ontology clusters.....	9
4. Comparison of the hyponymy structures.....	15
4.1. General properties of the ILI-graphs .....	17
4.2. Comparison of the ILI-graphs with WordNet1.5 .....	20
5. Conclusions .....	23
References.....	24
Appendix I Projection of complete chains on the German, French, Estonian and Czech wordnets.....	25
Appendix II Projection of partial chains on the German, French, Estonian and Czech wordnets.....	27
Appendix III Partial chains with 1 gap projected over the German, French, Estonian and Czech wordnet .....	31

## List of Tables

Table 1: Intersection of ILI-references in English (WN), German (DE), French (FR), Estonian (EE) and Czech (CZ)...	8
Table 2: Nominal Synsets clustered as 1stOrder Concepts .....	11
Table 3: Nominal Synsets clustered as 2ndOrder Concepts .....	12
Table 4: Verbal Synsets clustered as 2ndOrder Concepts .....	13
Table 5: Nominal Synsets clustered as 3rdOrder Concepts.....	13
Table 6: German, French, Estonian and Czech Nouns clustered over the WordNet1.5 Lexicographer's file codes .....	14
Table 7: German (DE), French (FR), Estonian (EE) and Czech (CZ) Verbs clustered over the WordNet1.5 Lexicographer's file codes .....	14
Table 8: ILI chains for verbs .....	18
Table 9: ILI chains for verbs .....	18
Table 10: Frequencies and ratios of noun chains / length /language.....	19
Table 11: Frequencies and ratios of verb chains / length /language.....	19
Table 12: Coverage of complete noun chains projected over WN1.5 structure.....	20
Table 13: Coverage of complete verb chains projected over WN1.5 structure .....	20
Table 14: Coverage of partial noun chains of NODES projected over WN1.5 structure .....	20
Table 15: Coverage of partial noun chains of EDGES projected over WN1.5 structure.....	21
Table 16: Coverage of partial VERB chains of NODES projected over WN1.5 structure.....	21
Table 17: Coverage of partial VERB chains of EDGES projected over WN1.5 structure.....	21
Table 18: Coverage of partial NOUN chains of EDGES with 1 gap projected over WN1.5 structure .....	22
Table 19: Coverage of partial VERB chains of NODES with 1 gap projected over WN1.5 structure.....	22
Table 20: Coverage of partial VERB chains of EDGES with 1 gap projected over WN1.5 structure .....	23
Table 21: Coverage of complete noun chains projected over German wordnet structure .....	25
Table 22: Coverage of complete verb chains projected over German wordnet structure.....	25
Table 23: Coverage of complete noun chains projected over French wordnet structure.....	25
Table 24: Coverage of complete verb chains projected over French wordnet structure.....	25
Table 25: Coverage of complete noun chains projected over Estonian wordnet structure.....	25
Table 26: Coverage of complete verb chains projected over Estonian wordnet structure .....	26
Table 27: Coverage of complete noun chains projected over Czech wordnet structure .....	26
Table 28: Coverage of partial noun chains of NODES projected over German wordnet structure.....	27
Table 29: Coverage of partial noun chains of EDGES projected over German wordnet structure .....	27
Table 30: Coverage of partial VERB chains of NODES projected over German wordnet structure .....	27
Table 31: Coverage of partial VERB chains of EDGES projected over German wordnet structure.....	27
Table 32: Coverage of partial noun chains of NODES projected over French wordnet structure .....	28
Table 33: Coverage of partial noun chains of EDGES projected over French wordnet structure.....	28
Table 34: Coverage of partial verb chains of NODES projected over French wordnet structure .....	28
Table 35: Coverage of partial verb chains of EDGES projected over French wordnet structure.....	28
Table 36: Coverage of partial noun chains of NODES projected over Estonian wordnet structure.....	29
Table 37: Coverage of partial noun chains of EDGES projected over Estonian wordnet structure .....	29
Table 38: Coverage of partial VERB chains of NODES projected over Estonian wordnet structure .....	29
Table 39: Coverage of partial VERB chains of EDGES projected over Estonian wordnet structure.....	29
Table 40: Coverage of partial noun chains of NODES projected over Czech wordnet structure .....	30
Table 41: Coverage of partial noun chains of EDGES projected over Czech wordnet structure.....	30
Table 42: Coverage of partial noun chains of NODES with 1 gap projected over German wordnet structure.....	31
Table 43: Coverage of partial noun chains of EDGES with 1 gap projected over German wordnet structure .....	31
Table 44: Coverage of partial VERB chains of NODES with 1 gap projected over German wordnet structure .....	31
Table 45: Coverage of partial VERB chains of EDGES with 1 gap projected over German wordnet structure.....	31
Table 46: Coverage of partial noun chains of NODES with 1 gap projected over French wordnet structure .....	31
Table 47: Coverage of partial noun chains of EDGES with 1 gap projected over French wordnet structure.....	32
Table 48: Coverage of partial verb chains of NODES with 1 gap projected over French wordnet structure .....	32
Table 49: Coverage of partial verb chains of EDGES with 1 gap projected over French wordnet structure.....	32
Table 50: Coverage of partial noun chains of NODES with 1 gap projected over Estonian wordnet structure.....	32
Table 51: Coverage of partial noun chains of EDGES with 1 gap projected over Estonian wordnet structure .....	32
Table 52: Coverage of partial VERB chains of NODES with 1 gap projected over Estonian wordnet structure .....	33
Table 53: Coverage of partial VERB chains of EDGES with 1 gap projected over Estonian wordnet structure.....	33
Table 54: Coverage of partial noun chains of NODES with 1 gap projected over Czech wordnet structure .....	33
Table 55: Coverage of partial noun chains of EDGES with 1 gap projected over Czech wordnet structure.....	33



## 1. Introduction

This deliverable describes the comparison of the final wordnets for German, French, Estonian and Czech. The comparison of the wordnets is based on the equivalence relations to the Inter-Lingual-Index in each wordnet. The list of ILI-records associated with the local synsets can be seen as a language-neutral representation of the wordnets in different languages. Three types of comparison are carried out:

- comparison of the intersection of the associated ILI-records (carried out by the University of Amsterdam)
- distribution of the associated ILI-records over the different top-ontology clusters (carried out by the University of Amsterdam)
- comparison of the hyponymy relations in the wordnets, projected on the associated ILI-records (carried out by the University Politecnica de Catalunya)

## 2. Intersection of the associated ILI-records

The size of each wordnet is between 7,5K and 22K synsets. For comparison, WordNet1.5 has a size of about 80K synsets for nouns and verbs. Not all synsets have an equivalence relation to the ILI. In other cases, different synsets refer to the same ILI-record or single synsets are linked to multiple ILI-records. Finally, local synsets may be linked to an ILI-record by complex equivalence relations (e.g. EQ\_NEAR\_SYNONYM, EQ\_HAS\_HYPERONYM, EQ\_HAS\_MERONYM, EQ\_ROLE) or to ILI-records with a different part of speech. The number of ILI-record references in a wordnet therefore only weakly correlates with the actual size and coverage of the wordnet. Nevertheless, we can state that all the ILI-records are somehow associated to a local synset and that the concept is somehow incorporated in the lexicalization of the language concerned, albeit via multiple and complex equivalence relations.

More practically, the intersection of associated ILI-records indicates the extent to which the wordnets can be used for cross-language retrieval or mapping. If only the ILI-records are considered that are linked by a simple EQ\_SYNONYM RELATION, the intersection would represent overlap in a very strict sense. Here we took all the associated ILI-records, regardless of the type of equivalence link, which indicates the maximal overlap possible. For retrieval purposes, such a more global matching may be more useful.

Table-1 gives an overview of the number of ILI-records referred to in each wordnet and the intersection between them. The figures are differentiated for nouns and verbs. The first column gives the absolute numbers, the second column gives the percentage of all ILI-records occurring in the union of all 4 resources (including WordNet1.5), the third column gives the percentage of the ILI-references occurring in the union of the German, French, Estonian and Czech wordnet only (which is 37,9% of all the nouns in WordNet15 and 55,5% of all the verbs).

Table 1: Intersection of ILL-references in English (WN), German (DE), French (FR), Estonian (EE) and Czech (CZ)

	Nouns			Verbs		
	Total	62780	23832	Total	12215	6791
	frequency	% of $\cup$ WN,DE,FR,EE,CZ)	% of $\cup$ DE,FR,EE,CZ)	frequency	% of $\cup$ (WN,DE,FR,EE,CZ)	% of $\cup$ (DE,FR,EE,CZ)
<b>DE</b>	8703	13,87%	36,52%	3993	32,64%	58,80%
<b>FR</b>	17528	27,94%	73,55%	4892	39,99%	72,04%
<b>EE</b>	4255	6,78%	17,85%	2130	17,41%	31,37%
<b>CZ</b>	9727	15,51%	40,81%	3097	25,31%	45,60%
$\cap$ (DE, FR)	5336	8,51%	22,39%	2054	16,79%	30,25%
$\cap$ (DE, EE)	2165	3,45%	9,08%	1195	9,77%	17,60%
$\cap$ (DE, CZ)	3523	5,62%	14,78%	1576	12,88%	23,21%
$\cap$ (FR, EE)	3178	5,07%	13,34%	1495	12,22%	22,01%
$\cap$ (FR, CZ)	6710	10,70%	28,16%	2174	17,77%	32,01%
$\cap$ (EE, CZ)	2192	3,49%	9,20%	1194	9,76%	17,58%
$\cap$ (DE, FR, EE)	1847	2,94%	7,75%	936	7,65%	13,78%
$\cap$ (DE, EE, CZ)	1377	2,20%	5,78%	797	6,51%	11,74%
$\cap$ (FR, EE, CZ)	1866	2,97%	7,83%	964	7,88%	14,20%
$\cap$ (FR, DE, CZ)	2875	4,58%	12,06%	1236	10,10%	18,20%
$\cap$ (DE,FR,EE,CZ)	1242	1,98%	5,21%	666	5,44%	9,81%

We see here that the intersection of all the 5 languages (**DE,FR,EE,CZ** and **WordNet1.5**) is low compared to the total union: 5,21% for nouns and 9,81% for verbs. The smallest set represents the maximal intersection, which is 4,255 for the Estonian nouns and 2,130 for the Estonian verbs. Compared to this, the total intersection is about 30% of the maximal possible intersection. Intersection of 3 and 2 languages compared to WordNet15 are higher. Because of the larger contribution of French, we see that the intersections with French are a bit higher than the others.

If we look at the intersection for 4 languages, the results are a bit lower than for EuroWordNet-1 (Dutch, Spanish, Italian and WordNet1.5). For EuroWordNet-1, the intersection for 3 languages was 23,8% for nouns and 21,9% for verbs. One cause for this is the fact that the size of the wordnets is much smaller, this reduces the chance that similar concepts are selected.



### 3. The distribution of the associated ILI-records over the top-ontology clusters

As explained in D014D015 (Vossen et al. 1998), the wordnets are built top-down starting with the Base Concepts. Each site is free to include different lexicalizations patterns when extending the vocabulary from the Base Concepts down. Still, to get an idea of the conceptual distribution of this extension we also measure the progress of the wordnets relative to the EuroWordNet Top Ontology (see Figure 1), which represents the diversity of Base Concepts that have been selected (for an explanation of the Top Ontology see Rodriquez et al 1998 and Vossen 1999). For this purpose, AMS implemented an inheritance mechanism that derives the Top Concepts from hyperonyms in WordNet1.5. By loading ILI-equivalences of the German, French, Czech and Estonian wordnets in the Amsterdam lexical database (ALS), it is possible to collect the Top Concepts that apply to these equivalences via hyponymy-inheritance in WordNet1.5. By applying this to all the equivalences, it is possible to quantify the coverage per top concept. Note that this measurement depends on the quality and quantity of the equivalence relations. Not all synsets have a (correct) equivalent relation. Furthermore, it may be that the hyponymy relations in the local wordnets are different, but according to this procedure they will all be classified by the same hyponymy-chains in WN1.5. This method thus gives a rough indication of the conceptual coverage.

The Top Ontology is divided in 3 main parts:

- 1stOrderEntities (nouns): concrete things
- 2ndOrderEntities (nouns, verbs and adjectives): states, events, processes, relations and properties
- 3rdOrderEntities (nouns): idea, knowledge, propositions

Top <sup>0</sup>	
1stOrderEntity <sup>1</sup>	2ndOrderEntity <sup>0</sup>
<b>Origin<sup>0</sup></b> Natural <sup>21</sup> Living <sup>30</sup> Plant <sup>18</sup> Human <sup>106</sup> Creature <sup>2</sup> Anima <sup>123</sup> Artifact <sup>144</sup> <b>Form<sup>0</sup></b> Substance <sup>32</sup> Solid <sup>63</sup> Liquid <sup>13</sup> Gas <sup>1</sup> Object1 <sup>62</sup> <b>Composition<sup>0</sup></b> Part <sup>86</sup> Group <sup>63</sup> <b>Function<sup>55</sup></b> Vehicle <sup>8</sup> Representation <sup>12</sup> MoneyRepresentation <sup>10</sup> LanguageRepresentation <sup>34</sup> ImageRepresentation <sup>9</sup> Software <sup>4</sup> Place <sup>45</sup> Occupation <sup>23</sup> Instrument <sup>18</sup> Garment <sup>3</sup> Furniture <sup>6</sup> Covering <sup>8</sup> Container <sup>12</sup> Comestible <sup>32</sup> Building <sup>13</sup>	<b>SituationType<sup>6</sup></b> Dynamic <sup>134</sup> BoundedEvent <sup>183</sup> UnboundedEvent <sup>48</sup> Static <sup>28</sup> Property <sup>61</sup> Relation <sup>38</sup> <b>SituationComponent<sup>0</sup></b> Cause <sup>67</sup> Agentive <sup>170</sup> Phenomenal <sup>17</sup> Stimulating <sup>25</sup> Communication <sup>50</sup> Condition <sup>62</sup> Existence <sup>27</sup> Experience <sup>43</sup> Location <sup>76</sup> Manner <sup>21</sup> Mental <sup>90</sup> Modal <sup>10</sup> Physical <sup>140</sup> Possession <sup>23</sup> Purpose <sup>137</sup> Quantity <sup>39</sup> Social <sup>102</sup> Time <sup>24</sup> Usage <sup>8</sup>
3rdOrderEntity <sup>33</sup>	

Figure 1: The EuroWordNet Top-Ontology

The results are given in the next tables, where nouns are divided into separate tables for 1st, 2nd and 3rdOrder Entities, and the verbs listed in one table of 2ndOrderEntities. It should be noted that we do not quantify the number of *synsets* but the number of *Top-Concept assignments* or Top-Concept tokens. Due to inheritance and multiple Top-Concept assignments, most synsets get several Top-Concepts. A Top-Concept is however only assigned once if it is derived via multiple paths or nodes.

In Table 2, the results are given for the 1st Order Entities. The first column lists the 1stOrder Top-Concepts. The next column gives the number of Top-Concept tokens or assignments in WordNet1.5: either directly or indirectly (via a hyperonym chain). The 3rd column gives the percentages of the total clusters in WordNet1.5. The 1st column of each wordnet gives the same TC-clustering based on the TC-inheritance in WordNet1.5 of the ILI-records representing the local

wordnet synsets. The next column gives the percentage of the total set of 1stOrder nouns covered by each wordnet and the 4th column for German (DE), French (FR), Estonian (EE) and Czech (CZ) gives the percentage of the corresponding TC clusters in WordNet1.5.

Table 2: Nominal Synsets clustered as 1stOrder Concepts

	WN15			DE			FR			EE			CZ		
Top-Concept	TC-Tokens	%of wn	TC-Tokens	% of de	%of wn	TC-Tokens	%of fr	%of wn	TC-Tokens	%of ee	%of wn	TC-Tokens	%of cz	%of wn	
Animal	14068	3.99%	607	2,42%	4,31%	633	1,48%	4,50%	115	1,12%	0,82%	418	1,37%	2,97%	
Artifact	19562	5.55%	1737	6,94%	8,88%	4772	11,1%	24,39%	1086	10,6%	5,55%	2709	8,88%	13,85%	
Building	1022	0.29%	113	0,45%	11,06%	251	0,59%	24,56%	45	0,44%	4,40%	194	0,64%	18,98%	
Comestible	3377	0.96%	629	2,51%	18,63%	831	1,95%	24,61%	174	1,70%	5,15%	295	0,97%	8,74%	
Container	1725	0.49%	140	0,56%	8,12%	295	0,69%	17,10%	82	0,80%	4,75%	191	0,63%	11,07%	
Covering	2030	0.58%	187	0,75%	9,21%	390	0,91%	19,21%	83	0,81%	4,09%	301	0,99%	14,83%	
Creature	664	0.19%	19	0,08%	2,86%	41	0,10%	6,17%	1	0,01%	0,15%	18	0,06%	2,71%	
Function	34081	9.68%	3441	12,1%	10,10%	7456	14,9%	21,88%	1640	13,9%	4,81%	4374	14,3%	12,83%	
Furniture	298	0.08%	48	0,19%	16,11%	76	0,18%	25,50%	27	0,26%	9,06%	38	0,12%	12,75%	
Garment	756	0.21%	101	0,40%	13,36%	168	0,39%	22,22%	27	0,26%	3,57%	107	0,35%	14,15%	
Gas	93	0.03%	15	0,06%	16,13%	20	0,05%	21,51%	14	0,14%	15,05%	13	0,04%	13,98%	
Group	27805	7.90%	1155	4,61%	4,15%	1282	3,00%	4,61%	334	3,26%	1,20%	847	2,78%	3,05%	
Human	11543	3.28%	1099	4,39%	9,52%	2389	5,59%	20,70%	448	4,37%	3,88%	1647	5,40%	14,27%	
ImageRepresentation	780	0.22%	54	0,22%	6,92%	159	0,37%	20,38%	49	0,48%	6,28%	106	0,35%	13,59%	
Instrument	7036	2.00%	721	2,88%	10,25%	1810	4,24%	25,72%	495	4,83%	7,04%	983	3,22%	13,97%	
LanguageRepresentation	2844	0.81%	150	0,60%	5,27%	546	1,28%	19,20%	172	1,68%	6,05%	406	1,33%	14,28%	
Liquid	1629	0.46%	175	0,70%	10,74%	278	0,65%	17,07%	83	0,81%	5,10%	96	0,31%	5,89%	
Living	47104	13.37%	2914	11,6%	6,19%	3864	9,04%	8,20%	799	7,80%	1,70%	2665	8,74%	5,66%	
MoneyRepresentation	372	0.11%	26	0,10%	6,99%	81	0,19%	21,77%	24	0,23%	6,45%	56	0,18%	15,05%	
Natural	68370	19.41%	5249	20,9%	7,68%	8624	20,1%	12,61%	2126	20,7%	3,11%	5548	18,1%	8,11%	
Object	48162	13.68%	4570	18,2%	9,49%	7990	18,7%	16,59%	1964	19,1%	4,08%	5146	16,8%	10,68%	
Occupation	2059	0.58%	203	0,81%	9,86%	465	1,09%	22,58%	93	0,91%	4,52%	332	1,09%	16,12%	
Part	12083	3.43%	1309	5,23%	10,83%	2250	5,27%	18,62%	568	5,54%	4,70%	1022	3,35%	8,46%	
Place	5281	1.50%	416	1,66%	7,88%	1093	2,56%	20,70%	198	1,93%	3,75%	471	1,54%	8,92%	
Plant	18874	5.36%	853	3,41%	4,52%	394	0,92%	2,09%	117	1,14%	0,62%	339	1,11%	1,80%	
Representation	934	0.27%	56	0,22%	6,00%	278	0,65%	29,76%	63	0,61%	6,75%	182	0,60%	19,49%	
Software	201	0.06%	30	0,12%	14,93%	44	0,10%	21,89%	18	0,18%	8,96%	23	0,08%	11,44%	
Solid	6319	1.79%	703	2,81%	11,13%	985	2,31%	15,59%	300	2,93%	4,75%	667	2,19%	10,56%	
Substance	12365	3.51%	1464	5,85%	11,84%	2127	4,98%	17,20%	618	6,03%	5,00%	1216	3,99%	9,83%	
Vehicle	747	0.21%	92	0,37%	12,32%	157	0,37%	21,02%	39	0,38%	5,22%	91	0,30%	12,18%	
Total	352184		25041		7,11%	42724		12,13%	10246		2,91%	30501		8,66%	

If the wordnets are equally balanced then the relative percentages of the wordnets should be the same, even if the total size of the wordnets is different. When a particular percentage is significantly lower than the other wordnets it means that this wordnet is not balanced in this domain. If WordNet1.5 is used as a comparison, the percentage of the 3rd column should be about 10% to 15%, since the aimed total size of the wordnets is about 1/6 of WordNet1.5. Furthermore, we should realize that these clusterings are based on the ILI-equivalences linked to the synsets. If no equivalences are given, we cannot derive Top-Concept assignments for this synset via WN15.

A known fact is that some areas such as Animal and Plant are very difficult to match because

WordNet1.5 contains a lot of expert terminology in these particular domains. Also Group is a field which is rather big in WordNet1.5. Coverage of these is less for all wordnets in EuroWordNet1 and EuroWordNet2. Some specific lower coverage is shown for LanguageRepresentation and Representation in the German wordnet, and Comestible in Estonian and Czech.

The next two tables show the distribution for nouns and verbs that are classified as 2ndOrderEntities according to the WordNet1.5 hyponymy chains.

Table 3: Nominal Synsets clustered as 2ndOrder Concepts

	WN15		DE			FR			EE			CZ		
Top-Concept	TC-Tokens	%of wn	TC-Tokens	% of de	%of wn	TC-Tokens	%of fr	%of wn	TC-Tokens	%of ee	%of wn	TC-Tokens	%of cz	%of wn
Agentive	12255	6.84%	763	5,71%	6,23%	2852	7,17%	23,27%	630	6,19%	5,14%	1467	7,00%	11,97%
BoundedEvent	8142	4.55%	552	4,13%	6,78%	1852	4,66%	22,75%	545	5,36%	6,69%	1033	4,93%	12,69%
Cause	15458	8.63%	1027	7,69%	6,64%	3512	8,83%	22,72%	862	8,47%	5,58%	1829	8,73%	11,83%
Communication	7097	3.96%	466	3,49%	6,57%	1552	3,90%	21,87%	439	4,31%	6,19%	1050	5,01%	14,79%
Condition	3951	2.21%	346	2,59%	8,76%	931	2,34%	23,56%	205	2,01%	5,19%	409	1,95%	10,35%
Dynamic	20026	11.18%	1523	11,40%	7,61%	4628	11,6%	23,11%	1150	11,3%	5,74%	2407	11,4%	12,02%
Existence	330	0.18%	9	0,07%	2,73%	86	0,22%	26,06%	20	0,20%	6,06%	52	0,25%	15,76%
Experience	6862	3.83%	638	4,77%	9,30%	1556	3,91%	22,68%	383	3,76%	5,58%	782	3,73%	11,40%
Location	1536	0.96%	89	0,67%	5,79%	388	0,98%	25,26%	95	0,93%	6,18%	176	0,84%	11,46%
Manner	934	0.52%	57	0,43%	6,10%	222	0,56%	23,77%	54	0,53%	5,78%	94	0,45%	10,06%
Mental	10444	5.83%	821	6,14%	7,86%	2358	5,93%	22,58%	604	5,94%	5,78%	1415	6,75%	13,55%
Modal	542	0.30%	49	0,37%	9,04%	122	0,31%	22,51%	22	0,22%	4,06%	57	0,27%	10,52%
Phenomenal	2132	1.19%	181	1,35%	8,49%	407	1,02%	19,09%	96	0,94%	4,50%	191	0,91%	8,96%
Physical	8066	4.50%	661	4,95%	8,19%	1717	4,32%	21,29%	521	5,12%	6,46%	874	4,17%	10,84%
Possession	1411	0.79%	94	0,70%	6,66%	308	0,77%	21,83%	56	0,55%	3,97%	165	0,79%	11,69%
Property	12336	6.89%	1095	8,19%	8,88%	3112	7,83%	25,23%	663	6,52%	5,37%	1283	6,12%	10,40%
Purpose	15275	8.53%	979	7,33%	6,41%	3149	7,92%	20,62%	770	7,57%	5,04%	1825	8,71%	11,95%
Quantity	3864	2.16%	377	2,82%	9,76%	725	1,82%	18,76%	266	2,61%	6,88%	263	1,25%	6,81%
Relation	6822	3.81%	439	3,29%	6,44%	1438	3,62%	21,08%	434	4,27%	6,36%	937	4,47%	13,73%
Social	12024	6.71%	795	5,95%	6,61%	2448	6,16%	20,36%	651	6,40%	5,41%	1461	6,97%	12,15%
Static	21365	11.93%	1770	13,25%	8,28%	4927	12,3%	23,06%	1194	11,7%	5,59%	2359	11,2%	11,04%
Stimulating	1119	0.62%	74	0,55%	6,61%	216	0,54%	19,30%	114	1,12%	10,19%	145	0,69%	12,96%
Time	1444	0.81%	170	1,27%	11,77%	220	0,55%	15,24%	99	0,97%	6,86%	134	0,64%	9,28%
UnboundedEvent	4567	2.55%	306	2,29%	6,70%	922	2,32%	20,19%	258	2,54%	5,65%	509	2,43%	11,15%
Usage	1084	0.61%	82	0,61%	7,56%	108	0,27%	9,96%	44	0,43%	4,06%	41	0,20%	3,78%
Total	179086		13363		7,46%	39756		22,20%	10175		5,68%	20958		11,70%

As in EuroWordNet-1, coverage of 2ndOrderEntities is better. Here we see a relative lower coverage for Existence (German), Possession (Estonian), Quantity (Czech) and Usage (French and Czech).

Table 4: Verbal Synsets clustered as 2ndOrder Concepts

	WN15		DE			FR			EE			CZ		
Top-Concept	TC-Tokens	%of wn	TC-Tokens	% of de	%of wn	TC-Tokens	%of fr	%of wn	TC-Tokens	%of ee	%of wn	TC-Tokens	%of cz	%of wn
Agentive	8176	7.1%	1019	7,93%	12,46%	1636	7,37%	20,01%	597	7,05%	7,30%	932	6,87%	11,40%
BoundedEvent	10262	8.9%	1114	8,67%	10,86%	1898	8,55%	18,50%	747	8,83%	7,28%	1219	8,99%	11,88%
Cause	15261	13.2%	1714	13,3%	11,23%	3026	13,6%	19,83%	1134	13,4%	7,43%	1768	13,0%	11,59%
Communication	3969	3.4%	481	3,74%	12,12%	848	3,82%	21,37%	303	3,58%	7,63%	448	3,30%	11,29%
Condition	1730	1.5%	182	1,42%	10,52%	293	1,32%	16,94%	111	1,31%	6,42%	200	1,47%	11,56%
Dynamic	23487	20.4%	2562	19,9%	10,91%	4495	20,2%	19,14%	1782	21,0%	7,59%	2830	20,8%	12,05%
Existence	2296	2.0%	253	1,97%	11,02%	470	2,12%	20,47%	176	2,08%	7,67%	259	1,91%	11,28%
Experience	2067	1.8%	271	2,11%	13,11%	365	1,65%	17,66%	176	2,08%	8,51%	244	1,80%	11,80%
Location	8184	7.1%	737	5,74%	9,01%	1536	6,92%	18,77%	587	6,94%	7,17%	1032	7,61%	12,61%
Manner	350	0.3%	31	0,24%	8,86%	82	0,37%	23,43%	26	0,31%	7,43%	29	0,21%	8,29%
Mental	3048	2.6%	383	2,98%	12,57%	511	2,30%	16,77%	213	2,52%	6,99%	319	2,35%	10,47%
Modal	101	0.1%	12	0,09%	11,88%	13	0,06%	12,87%	5	0,06%	4,95%	13	0,10%	12,87%
Phenomenal	129	0.1%	22	0,17%	17,05%	9	0,04%	6,98%	11	0,13%	8,53%	16	0,12%	12,40%
Physical	11642	10.1%	1192	9,28%	10,24%	2169	9,78%	18,63%	904	10,6%	7,76%	1427	10,5%	12,26%
Possession	1968	1.7%	254	1,98%	12,91%	372	1,68%	18,90%	116	1,37%	5,89%	218	1,61%	11,08%
Property	504	0.4%	72	0,56%	14,29%	105	0,47%	20,83%	30	0,35%	5,95%	59	0,43%	11,71%
Purpose	4436	3.8%	575	4,48%	12,96%	914	4,12%	20,60%	304	3,59%	6,85%	509	3,75%	11,47%
Quantity	690	0.6%	71	0,55%	10,29%	158	0,71%	22,90%	67	0,79%	9,71%	100	0,74%	14,49%
Relation	960	0.8%	111	0,86%	11,56%	170	0,77%	17,71%	56	0,66%	5,83%	109	0,80%	11,35%
Social	5706	4.9%	647	5,04%	11,34%	1077	4,85%	18,87%	361	4,27%	6,33%	615	4,53%	10,78%
Static	6217	5.4%	592	4,61%	9,52%	1071	4,83%	17,23%	404	4,77%	6,50%	690	5,09%	11,10%
Stimulating	878	0.8%	123	0,96%	14,01%	175	0,79%	19,93%	98	1,16%	11,16%	99	0,73%	11,28%
Time	98	0.1%	13	0,10%	13,27%	25	0,11%	25,51%	5	0,06%	5,10%	10	0,07%	10,20%
UnboundedEvent	2536	2.2%	345	2,69%	13,60%	659	2,97%	25,99%	216	2,55%	8,52%	343	2,53%	13,53%
Usage	646	0.6%	71	0,55%	10,99%	111	0,50%	17,18%	35	0,41%	5,42%	77	0,57%	11,92%
Total	115341		12847		11,14%	22188		19,24%	8464		7,34%	13565		11,76%

The 2ndOrder verbs are well covered too. Slightly less concepts are covered for Manner (German), Modal (Estonian) and Phenomenal (French).

Finally, the next table gives the nominal synsets classified as 3rdOrderEntities, where the percentages give the proportion of the set in WordNet1.5. Here we see that Czech and French have significantly more 3rdOrderEntities than the others, and also more than expected, given the total size of the wordnets. These percentages are even as high as in EuroWordNet-1.

Table 5: Nominal Synsets clustered as 3rdOrder Concepts

	WN15	DE		FR		EE		CZ	
	TC-Tokens	TC-Tokens	% of wn	TC-Tokens	% of wn	TC-Tokens	% of wn	TC-Tokens	% of wn
3rdOrderEntity	8059	561	6,96%	1714	21,27%	483	5,99%	1169	14,51%

Since we also added the WordNet1.5 lexicographer's file codes to the database it is also possible to measure the subsets with respect to that classification. This is shown in the next tables:

Table 6: German, French, Estonian and Czech Nouns clustered over the WordNet1.5 Lexicographer's file codes

Lexicographer's file code	WN15		DE		FR		EE		CZ	
	TC-Tokens	% of wn	TC-Tokens	% wn	TC-Tokens	% wn	TC-Tokens	% wn	TC-Tokens	% wn
4 noun.act	8582	6.83%	449	5,23%	1922	22,40%	397	4,63%	937	10,92%
5 noun.animal	13803	10.99%	554	4,01%	571	4,14%	101	0,73%	377	2,73%
6 noun.artifact	14994	11.94%	1350	9,00%	3527	23,52%	850	5,67%	2228	14,86%
7 noun.attribute	4741	3.78%	320	6,75%	1209	25,50%	200	4,22%	402	8,48%
8 noun.body	2900	2.31%	350	12,07%	432	14,90%	103	3,55%	275	9,48%
9 noun.cognition	3997	3.18%	374	9,36%	910	22,77%	246	6,15%	535	13,39%
10 noun.communication	6819	5.43%	431	6,32%	1484	21,76%	413	6,06%	1009	14,80%
11 noun.event	1389	1.11%	106	7,63%	352	25,34%	110	7,92%	207	14,90%
12 noun.feeling	758	0.60%	79	10,42%	210	27,70%	29	3,83%	61	8,05%
13 noun.food	3352	2.67%	623	18,59%	821	24,49%	166	4,95%	288	8,59%
14 noun.group	13728	10.93%	342	2,49%	667	4,86%	170	1,24%	402	2,93%
15 noun.location	3231	2.57%	267	8,26%	632	19,56%	96	2,97%	177	5,48%
16 noun.motive	53	0.04%	6	11,32%	11	20,75%		0,00%	2	3,77%
17 noun.object	4083	3.25%	429	10,51%	599	14,67%	144	3,53%	397	9,72%
18 noun.person	9356	7.45%	835	8,92%	1831	19,57%	303	3,24%	1313	14,03%
19 noun.phenomenon	751	0.60%	102	13,58%	158	21,04%	45	5,99%	76	10,12%
20 noun.plant	18536	14.76%	766	4,13%	334	1,80%	95	0,51%	307	1,66%
21 noun.possession	1240	0.99%	86	6,94%	272	21,94%	54	4,35%	144	11,61%
22 noun.process	1038	0.83%	32	3,08%	161	15,51%	38	3,66%	72	6,94%
23 noun.quantity	2021	1.61%	181	8,96%	374	18,51%	132	6,53%	66	3,27%
24 noun.relation	944	0.75%	90	9,53%	166	17,58%	64	6,78%	82	8,69%
25 noun.shape	633	0.50%	60	9,48%	131	20,70%	45	7,11%	76	12,01%
26 noun.state	3162	2.52%	309	9,77%	743	23,50%	115	3,64%	319	10,09%
27 noun.substance	4048	3.22%	542	13,39%	536	13,24%	279	6,89%	376	9,29%
28 noun.time	1427	1.14%	167	11,70%	210	14,72%	96	6,73%	132	9,25%
Total	125586		8850	7,05%	18263	14,54%	4291	3,42%	10260	8,17%

Table 7: German (DE), French (FR), Estonian (EE) and Czech (CZ) Verbs clustered over the WordNet1.5 Lexicographer's file codes

Lexicographer's file code	WN15		DE		FR		EE		CZ	
	TC-Tokens	% of wn	TC-Tokens	% wn	TC-Tokens	% wn	TC-Tokens	% wn	TC-Tokens	% wn
29 verb.body	1095	0.39%	142	12,97%	150	13,70%	92	8,40%	108	9,86%
30 verb.change	6379	2.24%	553	8,67%	1235	19,36%	470	7,37%	777	12,18%
31 verb.cognition	1986	0.70%	265	13,34%	338	17,02%	125	6,29%	209	10,52%
32 verb.communication	3569	1.26%	441	12,36%	776	21,74%	270	7,57%	409	11,46%
33 verb.competition	791	0.28%	65	8,22%	74	9,36%	28	3,54%	61	7,71%
34 verb.consumption	569	0.20%	65	11,42%	98	17,22%	33	5,80%	67	11,78%
35 verb.contact	4028	1.42%	314	7,80%	672	16,68%	256	6,36%	496	12,31%
36 verb.creation	1658	0.58%	175	10,55%	313	18,88%	134	8,08%	179	10,80%
37 verb.emotion	789	0.28%	90	11,41%	137	17,36%	72	9,13%	84	10,65%
38 verb.motion	3865	1.36%	345	8,93%	867	22,43%	310	8,02%	514	13,30%
39 verb.perception	870	0.31%	129	14,83%	168	19,31%	81	9,31%	106	12,18%
40 verb.possession	1815	0.64%	237	13,06%	342	18,84%	104	5,73%	196	10,80%
41 verb.social	4209	1.48%	499	11,86%	896	21,29%	295	7,01%	489	11,62%
42 verb.stative	1345	0.47%	163	12,12%	248	18,44%	103	7,66%	96	7,14%
43 verb.weather	117	0.04%	21	17,95%	9	7,69%	11	9,40%	16	13,68%
Total	284257		3504	10,59%	6323	19,11%	2384	7,21%	3807	11,51%

Similar tendencies can be observed here as well, e.g. Animal, Plant and Group come out as less well covered than in WordNet1.5. Some new observations are for nouns Location (Estonian and Czech), Motive (Estonian and Czech), Process (German) and for verbs, Competition (French, Estonian and Czech), Consumption (Estonian), Contact (German), Possession (Estonian), Stative (Czech), Weather (French). Most of these categories did not show up in the previous clustering, because of the difference in the classification.

#### 4. Comparison of the hyponymy structures

The previous comparisons only indicate the overlap in ILI-records and their conceptual clustering. To measure the compatibility of the hyponymy structures (which is the most important relation) we have to impose the relations on the ILI records as well.

For this comparison each site (DE, FR, EE, CZ) has generated sets of so-called ILI-chains for the nouns and verbs. These chains are based on the hyponymy relations but the original nouns and verbs are replaced by the ILI-records that are associated as eq\_synonym or eq\_near\_synonym. For example, the next list of Dutch senses is generated for "opstijgen" (take off) by recursively taking all the hyperonyms. When this chain is reversed we get the following list:

veranderen (change)  $\Leftarrow$  bewegen (move intransitive)  $\Leftarrow$  bewegen (move reflexive)  $\Leftarrow$  voortbewegen (move location)  $\Leftarrow$  verplaatsen (move from A to B)  $\Leftarrow$  stijgen (move to a higher position)  $\Leftarrow$  opstijgen (take off)

To be able to compare these chains, each word sense in the chain has been replaced by the ILI-records that are linked to these synsets which gives the following result:

00064108 01046072 01046072 01046072 01055491 01094615 00257753

This means that the Dutch equivalent to ILI record number 00064108 (veranderen) has the equivalent to ILI record number 01046072 (bewegen) as a hyponym and this one has the equivalent to ILI record number 01046072 as a hyponym, etc. It should be noted that the ILI-chains are in many ways partial representations of the wordnet structures. Not only may there be cases where nodes have no translations or complex equivalence relations, in which the original word is inserted in the chain, in other cases multiple translations have been assigned of which only one has been selected for generating the ILI-chains. If all combinations of chains were generated the number of chains would be too high. The compared graphs thus represent a simplification of the actual graphs.

The ILI-chains are imported as a graph and the sequences of other wordnets are compared to this graph by a special graph comparison tool developed by the University Politecnica de Catalunya. Two kinds of compatibility measurements can be applied to these chains with this tool:

- Edge-coverage of chains means that not only the synsets but also the hyponymy relations between them are covered by the different wordnets.
- Node-coverage of chains means that the synsets are covered but not necessarily the hyponymy relations. Perhaps another relation holds between the corresponding synsets or perhaps they are unrelated.

Consider, for instance, that languages L1 and L2 contains the following ILI chains:

L1: 1--2--3 & 1--4--5  
L2: 1--2 & 1--3--4--5

The chain 1--2--3--4--5 is node-covered by both L1 and L2 languages but is not completely edge-covered by any of them. There are, however, two sub-chains of length 3, one for each language, and 2 sub-chains of length 2, also one for each language, that have edge coverage. Note that node coverage can be the results of nodes that come from disjoint branches in the hierarchy. A language that covers all ILIs but has no hyponymy relations (L3: 1 & 2 & 3 & 4 & 5) will thus also have full node coverage.

Both measurements are important and can be used in different way. Of course edge-coverage is difficult to achieve (covering an edge implies covering the two related nodes and the relation between them -in the same direction-). A high degree of edge-covering overlap means that the overlapping concepts exist and are lexicalized in all the languages that overlap and that their structural (hyponym/hyperonym) relationships hold in the same way for such languages (in so far as they are adequately represented by the associated ILI-records). A lower level of edge-covering overlapping could indicate:

- a) incompleteness in covering the nodes (can be measured by node-coverage)
- b) incompleteness of relations in the language (can be measured by edge-coverage)
- c) A genuine difference between vocabularies of the languages or the classification

Complete overlapping of chains (either at edge or node level) is impossible due to the (huge) differences in size of the wordnets to be compared. However, complete compatibility with WordNet1.5 or any of the wordnets is not the goal in EuroWordNet. There are differences at the highest level of the hierarchy that are based on different insights or differences in lexicalization. For example, WordNet1.5 has 573 tops for verbs, whereas for example the Estonian wordnet has a more unified the verb hierarchy with 94 tops. In that case there can never be full compatibility. We have therefore used two additional measurements:

- Sub-sequences of N-length: simply chains of nodes/edges that exactly match a fragment of another chain.
- Sub-sequences of N-lengths with M gaps: chains of nodes/edges that match a fragment of another chain but failing to match M nodes of edges.

For example:

- Node sub-sequence of length 2:  
Sequence:  
00002728 00004865 05839075 06193747  
Sub-sequence:  
00004865 05839075
- Edge sub-sequence of length 2:  
Sequence:  
00002728 00004865 05839075 06193747  
Sub-sequence:  
00004865 05839075 06193747
- Node sub-sequence of length 3 with 1 gap:  
Sequence:  
00002728 00004865 05839075 06193747  
Sub-sequence:  
00004865 06193747
- Edge sub-sequence of length 4 with 2 gap:  
Sequence:



00002728 00004865 05839075 06193747 01137195  
 Sub-sequence:  
 00002728 00004865 06193747 01137195

Sub-sequences with 1 and 2 gaps are reported here. Although other cases can be computed in an easy way, they are less useful.

The procedure to extract the statistics consists of four steps:

1. One of the WNs is taken as a base. The set of chains is read and a graph structure (in fact a DAG) is built.
2. The other WNs are projected over this base. Possible cycles are not allowed. All the nodes are incorporated into the graph but only the compatible edges are added (i.e. the graph can be extended with additional nodes, some of the existing nodes can be marked as covered by the new language and some of the edges too, new edges can be added but only in the case they don't produce cycles).
3. The graph once completed is fully traversed in order to generate all the paths covering it (from tops to leaves). The set of paths is written into a file.
4. The file is queried in a variety of ways for extracting the statistics.

This procedure has been carried out 5 times, taking each wordnet as a starting point: WN1.5, German-WN, French-WN, Estonian-WN and Czech-WN (only the nouns). Next, we can query the database in a normal or verbose way. When using the verbose mode, not only the number but also the actual occurrences of the overlapping cases are extracted. Normal mode is used here for presenting the results and extracting some conclusions. The verbose mode has been used to select mismatches or uncovered ILI nodes and edges during the building of the wordnets.

In the next sections, we will represent the following quantitative data generated by the tool:

- 1) Individual level (data provided by each site without any cross comparison).
- 2) Degree of coverage of WN1.5.
- 3) Overlapping with the other sites.

The overlapping of the graphs across the different wordnet is given in the appendix. In the next section we will look the compatibility with WordNet1.5. Further details on the comparison can be found in D014D015 (Vossen et al. 1998).

#### ***4.1. General properties of the ILI-graphs***

The next tables give some general figures on the size and structure of the graphs. For Czech, no verbal ILI-chains have been received. A distinction is made between the tops, leaves, internal nodes, edges and chains:

- tops: end points without hyperonyms;
- leaves: end-points without hyponyms;
- internal-nodes: at least 1 hyponym and 1 hyperonym;
- edges: number of edges appearing in the sets, where each hyponymy connection represents an edge;
- chains: number of chains that can be generated from the edges;

Isolated ILI-records without hyponyms and hyperonyms are not considered by the program, since it tries to measure the compatibility of the relations.

Table 8: ILI chains for verbs

	<i>ILI nodes</i>	<i>Tops</i>	<i>Leaves</i>	<i>Internal Nodes</i>	<i>EDGES</i>	<i>CHAINS</i>
<b>WN15</b>	60557	11	47110	13436	61123	53467
<b>DE</b>	9357	554	7180	2224	10790	15341
<b>FR</b>	17527	22	11317	6191	17725	13401
<b>EE</b>	4801	23	3544	1282	5148	4542
<b>CZ</b>	9457	994	7135	1328	8498	7170

Table 9: ILI chains for verbs

	<i>ILI nodes</i>	<i>Tops</i>	<i>Leaves</i>	<i>Internal Nodes</i>	<i>EDGES</i>	<i>CHAINS</i>
<b>WN15</b>	11363	573	8446	2580	10816	8486
<b>DE</b>	3532	257	2899	1025	5525	8582
<b>FR</b>	4986	286	3533	1335	4713	3548
<b>EE</b>	2302	94	1716	596	2478	2109

If we look at the number of nodes, we see that the larger size of the French nouns is reflected in the number of ILI-nodes. The larger number of tops for German and Czech nouns may indicate less structure at the higher levels. The same holds for German and French verbs, although they have still less tops than WordNet1.5. At least this means that the complete wordnets are more difficult to access top-down.

The ratio of tops, leaves, and internal nodes tells us something about distribution of the nodes over different levels. Many leaves and few internal nodes indicates flat hierarchies, many internal nodes and relatively few leaves either indicates a deep or a tangled hierarchy. Except for German nouns and verbs, all the wordnets have less edges and chains than nodes. Here we see that the distributions are more or less the same across the wordnets. No extra-ordinary proportions.

Finally, a large proportion of chains relative to the number of nodes means a tangled hierarchy. This can be due to:

- multiple hyperonyms
- multiple translations
- large sets of synsets with the same translation

If the number of chains is extremely low, this indicates a lack of hyperonyms or translations. Since WordNet1.5 has an ideal mapping to the ILI (1:1) and it only occasionally incorporates multiple hyperonyms, we can expect that it represents a relatively ideal tree. The number of chains is a bit less than the number of nodes and we see that the French, Estonian and Czech wordnets have a similar proportion. The German wordnet has a bit more chains and edges, which is probably due to multiple hyperonyms. The other wordnets mostly must have single hyperonym relations.

In the case of a tangled structure, we can expect that the number of chains is bigger than the number of edges. The number of edges represents the number of hyponymy connections, but the number of chains represents the number of complete paths. In a tangled hierarchy, the same edges can occur in different chains. This is indeed the case for the German wordnet.

The next two tables present the number and % of noun and verb chains classified by length for each language.

Table 10: Frequencies and ratios of noun chains / length / language

	WN		DE		FR		EE		CZ	
	freq.	%	freq.	%	freq.	%	freq.	%	freq.	%
1			425	2.77	3	0.02	4	0.09		
2	33	0.06	459	2.99	28	0.21	63	1.39	2122	29.60
3	521	0.97	798	5.20	249	1.86	256	5.64	2652	36.99
4	2220	4.15	1602	10.44	908	6.78	448	9.86	1506	21.00
5	5664	10.59	2491	16.24	1913	14.28	818	18.01	668	9.32
6	12730	23.81	2626	17.12	2669	19.92	964	21.22	186	2.59
7	11741	21.96	2097	13.67	3101	23.14	656	14.44	35	0.49
8	8737	16.34	1523	9.93	2333	17.41	573	12.62	1	0.01
9	5940	11.11	1241	8.09	1171	8.74	433	9.53		
10	3305	6.18	849	5.53	700	5.22	235	5.17		
11	1400	2.62	620	4.04	185	1.38	82	1.81		
12	517	0.97	266	1.73	82	0.61	10	0.22		
13	364	0.68	93	0.61	41	0.31				
14	213	0.40	123	0.80	15	0.11				
15	75	0.14	97	0.63	3	0.02				
16	7	0.01	17	0.11						
17			14	0.09						
<b>Total</b>	53467	100	15341	100	13401	100	4542	100	7170	100
<b>Avg.</b>	7.19		6.50		6.85		6.39		3.20	

Table 11 Frequencies and ratios of verb chains / length / language

	WN		DE		FR		EE	
	freq.	%	freq.	%	freq.	%	freq.	%
1	236	2.78	128	1.49	168	4.74	4	0.09
2	1867	22.00	537	6.26	811	22.86	63	1.39
3	2530	29.81	1057	12.32	985	27.76	256	5.64
4	1959	23.09	1804	21.02	775	21.84	448	9.86
5	1029	12.13	1203	14.02	416	11.72	818	18.01
6	462	5.44	931	10.85	200	5.64	964	21.22
7	250	2.95	931	10.85	124	3.49	656	14.44
8	109	1.28	1504	17.53	51	1.44	573	12.62
9	32	0.38	386	4.50	13	0.37	433	9.53
10	10	0.12	93	1.08	4	0.11	235	5.17
11	2	0.02	8	0.09	1	0.03	82	1.81
12							10	0.22
<b>Total</b>	8486	100	8582	100	3548	100	4542	100
<b>Avg.</b>	3.58		5.39		3.54		6.39	

We see here that the German and French noun chains are most compatible in length with WordNet1.5. The Estonian chains are a bit shorter, although on average the difference is minimal. The Czech structures are very short, half the size. For the verbs, WN1.5 has shorter chains than the other wordnets. The Estonian wordnets are rather long given the small size. To some extent, this correlates with the smaller number of tops.

## 4.2. Comparison of the ILLI-graphs with WordNet1.5

The next tables account for the coverage of complete chains (at node and edge level) for nouns and verbs, projected over WN1.5. Projections over the other wordnets are listed in the Appendix.

Table 12: Coverage of complete noun chains projected over WN1.5 structure

	nodes (53467)		edges (53467)	
	frequency	%	frequency	%
DE	683	1.28	32	0.06
FR	10361	19.38	10354	19.37
EE	1040	1.95	394	0.74
CZ	218	0.41	10	0.02
$\cap(\text{DE,FR,EE})$	37	0.07	1	0.00
$\cap(\text{DE,FR,CZ})$	6	0.01	0	0.00
$\cap(\text{DE,EE,CZ})$	4	0.01	0	0.00
$\cap(\text{FR,EE,CZ})$	14	0.03	1	0.00
$\cap(\text{DE,FR,EE,CZ})$	3	0.01	0	0.00

Table 13: Coverage of complete verb chains projected over WN1.5 structure

	nodes (8486)		edges (8486)	
	frequency	%	frequency	%
DE	1016	11.97	184	2.17
FR	2714	31.98	2714	31.98
EE	412	4.86	184	2.17
$\cap(\text{DE,FR})$	584	6.88	74	0.87
$\cap(\text{DE,EE})$	154	1.81	41	0.48
$\cap(\text{FR,EE})$	234	2.76	97	1.14
$\cap(\text{DE,FR,EE})$	105	1.24	16	0.19

The figures presented in these tables are of rather limited use, since full coverage of the chains is rather difficult. It is therefore more important to look at the coverage of sub-chains of WN1.5 rather than the complete chains. The following four tables account for the overlap of partial chains (node vs. edge, noun vs. verb) projected over WN1.5 structure, for different lengths of the chains.

Table 14: Coverage of partial noun chains of NODES projected over WN1.5 structure

LEN	DE	FR	EE	CZ	$\cap(\text{DE,F,R,EE})$	$\cap(\text{DE,F,R,CZ})$	$\cap(\text{DE,E,E,CZ})$	$\cap(\text{FR,E,E,CZ})$	$\cap(\text{DE,FR,E,E,CZ})$	WN
1	53454	53467	53467	53156	53452	51585	51189	52950	51183	53467
2	47119	53438	47640	37466	40122	27608	24856	33615	24836	53467
3	31942	52991	44330	16125	27989	6611	5430	11209	5420	53434
4	19290	50851	36202	7838	14821	2122	1624	4499	1609	52913
5	10557	41779	24484	2974	6521	759	526	1390	525	50693
6	5294	30483	14541	1000	3395	88	20	402	20	45029
7	1583	19823	5897	225	510	9		69		32299
8	382	10893	2127	40	124			3		20558
9	85	5091	799	3	56					11821
10	33	2275	291		10					5881
11		1000	87							2576
12		457	6							1176
13		186								659
14		51								295
15		3								82

Table 15: Coverage of partial noun chains of EDGES projected over WN1.5 structure

LEN	DE	FR	EE	CZ	$\cap(DE,F,R,EE)$	$\cap(DE,F,R,CZ)$	$\cap(DE,E,E,CZ)$	$\cap(FR,E,E,CZ)$	$\cap(DE,FR,E,E,CZ)$	WN
1	36306	53438	44866	16484	27495	2566	884	6277	871	53467
2	11222	52991	36673	3999	8234	171	149	743	149	53434
3	5362	50851	28518	1283	4414	26	4	46	4	52913
4	554	41695	18154	237	250					50693
5	187	30425	9581	27	123					45029
6	5	19820	2954	2						32299
7		10891	1278							20558
8		5090	457							11821
9		2275	226							5881
10		1000	60							2576
11		457								1176
12		186								659
13		51								295
14		3								82

Table 16: Coverage of partial VERB chains of NODES projected over WN1.5 structure

LEN	DE	FR	EE	$\cap(DE,FR)$	$\cap(DE,EE)$	$\cap(FR,EE)$	$\cap(DE,FR,EE)$	WN
1	7958	8162	7417	7725	6979	7212	6853	8486
2	5203	6502	3677	4825	2815	3475	2733	8250
3	2695	4380	1162	2480	786	1092	775	6383
4	1265	2382	292	1189	171	270	165	3853
5	524	1152	53	483	12	50	12	1894
6	221	550	5	199		4		865
7	86	251	2	80		2		403
8	24	85		21				153
9	8	22		6				44
10	3	6		3				12
11		1						2

Table 17: Coverage of partial VERB chains of EDGES projected over WN1.5 structure

LEN	DE	FR	EE	$\cap(DE,FR)$	$\cap(DE,EE)$	$\cap(FR,EE)$	$\cap(DE,FR,EE)$	WN
1	2362	6493	2589	2203	695	2504	680	8250
2	370	4378	410	339	32	388	32	6383
3	78	2382	55	74		54		3853
4	1	1152	2			2		1894
5		550						865
6		251						403
7		85						153
8		22						44
9		6						12
10		1						2

The sub-sequences of node coverage more or less indicate the maximum coverage that is possible with the set of ILI-references that is given for each language. Sub-chains of length 1 are not interesting since the coverage can result from two unrelated sub-chains of length 1 projected from the wordnet on the WordNet1.5 graph. Node sub-chains of length 2 are perhaps not very meaningful either. The other tables with edge coverage than show how compatible the sub-sequences are in terms of the hyponymy relations. Edge coverage for longer sub-chains is extremely low.

The intersection of partial nodes for 3 languages is less than in EuroWordNet-1 (nouns 40636

length 2, 21089 length 3 and verbs 3,126 length 2, 1,072 length 3). This is partly due to the smaller size of the wordnets (about 50%). The same holds for edges. Remarkable is however the large amount of intersecting edges for German, French and Estonian nouns: 8,234 length 2 and 4414 length 3, compared to 1,113 length 2 and 113 length 3 in EuroWordNet-1!

The following tables then give the overlapping of partial chains with one gap (node vs. edge, noun vs. verb) projected over WN1.5 for different lengths of the chain. The Appendix gives the projections over the German, French, Estonian and Czech structures..

LEN	DE	FR	EE	CZ	$\cap(DE,F,R,EE)$	$\cap(DE,F,R,CZ)$	$\cap(DE,E,E,CZ)$	$\cap(FR,E,E,CZ)$	$\cap(DE,FR,E,E,CZ)$	WN
3	28891	101	7840	21997	25154	15180	12604	17360	12492	53434
4	24254	101	7707	16739	20021	8933	6767	12164	6725	52913
5	17833	101	7337	9162	12393	3220	2610	5672	2573	50693
6	11502	101	6125	3558	6887	527	284	1628	271	45029
7	6501	101	3600	1110	3227	166	128	349	125	32299
8	3166	100	2106	302	1431	13	1	104	1	20558
9	1027	16	694	32	395	7				11821
10	278	14	315	2	133					5881
11	40	3	34		20					2576
12	10	2	20		4					1176
13	2	1	4							659
14		1								295

Table 18: Coverage of partial NOUN chains of EDGES with 1 gap projected over WN1.5 structure

LEN	DE	FR	EE	CZ	$\cap(DE,F,R,EE)$	$\cap(DE,F,R,CZ)$	$\cap(DE,E,E,CZ)$	$\cap(FR,E,E,CZ)$	$\cap(DE,FR,E,E,CZ)$	WN
3	3831	59	2680	313	2881	0	0	11	0	52913
4	2575	59	1981	117	2123					50693
5	2062	59	1519	28	1789					45029
6	206	4	993	4	22					32299
7	18	1	315		3					20558
8	1	1	71							11821
9			17							5881
10			5							2576
11			1							1176

Table 19: Coverage of partial VERB chains of NODES with 1 gap projected over WN1.5 structure

LEN	DE	FR	EE	$\cap(DE,FR)$	$\cap(DE,EE)$	$\cap(FR,EE)$	$\cap(DE,FR,EE)$	WN
3	742	0	1488	625	1313	1375	1262	6383
4	407		878	344	695	822	677	3853
5	181		352	149	287	333	280	1894
6	98		104	84	58	96	53	865
7	47		18	39	5	16	5	403
8	18		1	15	1	1	1	153
9	2			2				44

Table 20: Coverage of partial VERB chains of EDGES with 1 gap projected over WN1.5 structure

LEN	DE	FR	EE	$\cap(DE,FR)$	$\cap(DE,EE)$	$\cap(FR,EE)$	$\cap(DE,FR,EE)$	WN
3	151	0	37	145	2	37	2	3853
4	39		3	37		2		1894
5	5			3				865

## 5. Conclusions

In this document we described the compatibility of the German, French, Estonian and Czech wordnet, especially compared to Wordnet1.5 and measured in terms of the ILI-references of their synsets. The first comparison involved the ILI-records that are referred to by the equivalence relations of the local wordnets to the ILI. The total intersection in ILI-references for all the 4 languages is about 5,21% for nouns and 9,81% for verbs respectively, where the maximal intersection is limited by the smallest set, which is 20% of the nouns and 30% of the verbs. If we look at the union of the ILI-references for the 3 languages, intersection is about 5,78% to 12,06% for nouns and from 11,74% to 18,20% for verbs. This is lower than for EuroWordNet-1.

The figures give the maximal matching across the 4 languages, regardless of the type of equivalence relation. The matching across language-pairs is higher: up to 28% for nouns and up to 32% for verbs. For cross-language retrieval this may be still a good basis, especially since it is possible to traverse the hierarchies in the local wordnet to get around mismatches in another language.

In addition, we looked at the distribution of these ILI-reference over the top-ontology. In general, these distributions are balanced across the wordnets. Relatively lower coverage has been measured for *plant*, *animal* and *group* nouns in all 3 wordnets, compared to WordNet1.5. Relatively higher coverage is achieved for abstract nouns and verbs. For some fields, slightly less coverage was measured.

The final comparison involved the hyponymy relations projected on the ILI-references, resulting in so-called ILI-chains. Hardly any overlap is measured in complete chains. This cannot be expected given the lower size compared to WordNet1.5 and the different choices at the top levels of the hierarchy. Lower figures in compatibility have been achieved for subsequences of nodes and edges than in EuroWordNet-1. This is partly due to the lower coverage. In the case of partial noun-edges for German, French and Estonian better figures were shown than for EuroWordNet-1.

## References

- Atserias J., S. Climent, J. Farreres, G. Rigau, H. Rodriguez,  
1997 *Combining Multiple Methods for the Automatic Construction of Multilingual WordNets*, Proceedings of Conference on Recent Advances on NLP. RANLP 97. Tzigov Chark, Bulgaria, 1997.
- Miller G.A, R. Beckwith, C. Fellbaum, D. Gross, and K.J. Miller,  
1990 *Introduction to WordNet: An On-line Lexical Database*, In: International Journal of Lexicography, Vol. 3, No.4, 235-244.
- Rodriquez, H., S. Climent, P. Vossen, L. Bloksma A. Roventini, F. Bertagna, A. Alonge, W. Peters,  
1998 The Top-Down Strategy for Building EuroWordNet: Vocabulary Coverage, Base Concepts and Top Ontology. In: Nancy Ide, Daniel Greenstein, Piek Vossen (eds), Special Issue on EuroWordNet. Computers and the Humanities, Volume 32, Nos. 2-3 1998. 117-152.
- Vossen, P., L. Bloksma, H. Rodriguez, S. Climent, N. Calzolari, A. Roventini, F. Bertagna, A. Alonge, W. Peters.  
1997 The EuroWordNet Base Concepts and Top Ontology. EuroWordNet (LE 4003) Deliverable D017D034D036. University of Amsterdam
- Vossen, P., L. Bloksma, S. Climent, M.A. Marti, G. Oreggioni, G. Escudero, G. Rigau, H. Rodriguez, C. Peters, A. Roventini, F. Bertagna, A. Alonge, W. Peters.  
1998 The Restructured Core wordnets in EuroWordNet: Subset1. EuroWordNet (LE 4003), Deliverable D014D015, University of Amsterdam.
- Vossen, P.  
1998 Vossen, P. Introduction to EuroWordNet. In: Nancy Ide, Daniel Greenstein, Piek Vossen (eds), Special Issue on EuroWordNet. Computers and the Humanities, Volume 32, Nos. 2-3 1998. 73-89.
- Vossen, P., L. Bloksma, S. Climent, M.A. Marti, m. Taule, J. Gonzalo, I. Chugur, F. Verdejo, G. Escudero, G. Rigau, H. Rodriguez, A. Alonge, F. Bertagna.  
1998 EuroWordNet Restructured Subset2 for Dutch, Spanish and Italian. EuroWordNet (LE 4003), Deliverable D027D028, University of Amsterdam.
- Vossen, P., L. Bloksma, S. Climent, M.A. Marti, m. Taule, J. Gonzalo, I. Chugur, F. Verdejo, G. Escudero, G. Rigau, H. Rodriguez, A. Alonge, F. Bertagna, Rita Marinelli, Adriana Roventini, Luca Tarasi.  
1999 Comparison of the Final Wordnets Dutch, Spanish and Italian, EuroWordNet (LE 4003), Deliverable D029D030, University of Amsterdam.



## Appendix I Projection of complete chains on the German, French, Estonian and Czech wordnets

Table 21: Coverage of complete noun chains projected over German wordnet structure

	<i>nodes</i>		<i>edges</i>	
	<i>frequency</i>	<i>%</i>	<i>frequency</i>	<i>%</i>
<b>FR</b>	2708	17.65	289	1.88
<b>EE</b>	483	3.15	80	0.52
<b>CZ</b>	359	2.34	140	0.91
$\cap(\text{FR,EE})$	423	2.76	59	0.38
$\cap(\text{FR,CZ})$	289	1.88	105	0.68
$\cap(\text{EE,CZ})$	94	0.61	42	0.27
$\cap(\text{FR,EE,CZ})$	83	0.54	37	0.24

Table 22: Coverage of complete verb chains projected over German wordnet structure

	<i>nodes</i>		<i>edges</i>	
	<i>frequency</i>	<i>%</i>	<i>frequency</i>	<i>%</i>
<b>FR</b>	1981	23.08	126	1.47
<b>EE</b>	533	6.21	68	0.79
$\cap(\text{FR,EE})$	422	4.92	44	0.51

Table 23: Coverage of complete noun chains projected over French wordnet structure

	<i>nodes</i>		<i>edges</i>	
	<i>frequency</i>	<i>%</i>	<i>frequency</i>	<i>%</i>
<b>DE</b>	469	3.50	18	0.13
<b>EE</b>	857	6.40	312	2.33
<b>CZ</b>	160	1.19	11	0.08
$\cap(\text{DE,EE})$	76	0.57	1	0.01
$\cap(\text{DE,CZ})$	14	0.10	1	0.01
$\cap(\text{EE,CZ})$	20	0.15	2	0.01
$\cap(\text{DE,EE,CZ})$	5	0.04	0	0.00

Table 24: Coverage of complete verb chains projected over French wordnet structure

	<i>nodes</i>		<i>edges</i>	
	<i>frequency</i>	<i>%</i>	<i>frequency</i>	<i>%</i>
<b>DE</b>	903	25.45	201	5.67
<b>EE</b>	413	11.64	206	5.81
$\cap(\text{DE,EE})$	211	5.95	64	1.80

Table 25: Coverage of complete noun chains projected over Estonian wordnet structure

	<i>nodes</i>		<i>edges</i>	
	<i>frequency</i>	<i>%</i>	<i>frequency</i>	<i>%</i>
<b>DE</b>	270	5.94	12	0.26
<b>FR</b>	2340	51.52	449	9.89
<b>CZ</b>	125	2.75	9	0.20
$\cap(\text{DE,FR})$	220	4.84	3	0.07
$\cap(\text{DE,CZ})$	26	0.57	1	0.02
$\cap(\text{FR,CZ})$	97	2.14	4	0.09
$\cap(\text{DE,FR,CZ})$	25	0.55	1	0.02

Table 26: Coverage of complete verb chains projected over Estonian wordnet structure

	<i>nodes</i>		<i>edges</i>	
	<i>frequency</i>	<i>(2109)</i>	<i>frequency</i>	<i>(2109)</i>
		<i>%</i>		<i>%</i>
<b>DE</b>	692	32.81	39	1.85
<b>FR</b>	1135	53.82	91	4.31
$\cap(\mathbf{DE,FR})$	521	24.70	14	0.66

Table 27: Coverage of complete noun chains projected over Czech wordnet structure

	<i>nodes</i>		<i>edges</i>	
	<i>frequency</i>	<i>(7170)</i>	<i>frequency</i>	<i>(7170)</i>
		<i>%</i>		<i>%</i>
<b>DE</b>	927	12.93	178	2.48
<b>FR</b>	4114	57.38	1108	15.45
<b>EE</b>	549	7.66	152	2.12
$\cap(\mathbf{DE,FR})$	613	8.55	101	1.41
$\cap(\mathbf{DE,EE})$	213	2.97	33	0.46
$\cap(\mathbf{FR,EE})$	382	5.33	89	1.24
$\cap(\mathbf{DE,FR,EE})$	152	2.12	25	0.35

## Appendix II Projection of partial chains on the German, French, Estonian and Czech wordnets

Table 28: Coverage of partial noun chains of NODES projected over German wordnet structure

LEN	FR	EE	CZ	$\cap(FR,EE)$	$\cap(FR,CZ)$	$\cap(EE,CZ)$	$\cap(FR,EE,CZ)$	DE
1	14939	14525	14512	14511	14413	13913	13883	15341
2	14050	12941	10038	12916	9868	8638	8601	14916
3	11780	8927	3948	8717	3786	2575	2509	14457
4	7428	4411	1068	4351	979	340	305	13659
5	4368	1342	278	1313	245	42	42	12057
6	1670	410	100	397	86	1	1	9566
7	623	122	21	115	19			6940
8	264	30	6	28	6			4843
9	74	4		2				3320
10	9							2079
11	2							1230

Table 29: Coverage of partial noun chains of EDGES projected over German wordnet structure

LEN	FR	EE	CZ	$\cap(FR,EE)$	$\cap(FR,CZ)$	$\cap(EE,CZ)$	$\cap(FR,EE,CZ)$	DE
1	10036	9000	1886	7997	980	504	393	14916
2	1786	1188	131	860	74	52	44	14457
3	550	307	8	240	7	2	2	13659
4	183	79		68				12057
5	55	34		34				9566
6	2							6940

Table 30: Coverage of partial VERB chains of NODES projected over German wordnet structure

LEN	FR	EE	$\cap(FR,EE)$	DE
1	8399	8033	7944	8582
2	7524	4997	4855	8454
3	4782	2108	1996	7917
4	2328	833	744	6860
5	1248	518	463	5056
6	718	255	235	3853
7	431	106	97	2922
8	168	35	28	1991
9	35	1	1	487
10	12			101
11	3			8

Table 31: Coverage of partial VERB chains of EDGES projected over German wordnet structure

LEN	FR	EE	$\cap(FR,EE)$	DE
1	2736	2058	1278	8454
2	557	52	5	7917
3	35			6860

Table 32: Coverage of partial noun chains of NODES projected over French wordnet structure

LEN	DE	EE	CZ	$\cap(DE,EE)$	$\cap(DE,CZ)$	$\cap(EE,CZ)$	$\cap(DE,EE,CZ)$	FR
1	13382	13392	13276	13374	12808	13204	12666	13401
2	11628	13185	10542	11084	7866	9117	6598	13398
3	8522	12085	5149	7166	2436	3304	1960	13370
4	5014	9367	2716	3561	844	1480	668	13121
5	2771	6316	1166	1837	346	561	264	12213
6	1285	3807	464	726	37	222	9	10300
7	468	1765	115	193	2	45		7631
8	134	813	21	62		3		4530
9	51	346	3	38				2197
10	22	119		8				1026
11		27						326
12		4						141

Table 33: Coverage of partial noun chains of EDGES projected over French wordnet structure

LEN	DE	EE	CZ	$\cap(DE,EE)$	$\cap(DE,CZ)$	$\cap(EE,CZ)$	$\cap(DE,EE,CZ)$	FR
1	8252	12286	5847	6986	1227	2194	268	13398
2	2251	9840	1501	1400	67	239	48	13370
3	551	7112	561	329	8	23	2	13121
4	141	4372	131	56				12213
5	37	2254	10	21				10300
6	2	900						7631
7		461						4530
8		178						2197
9		68						1026
10		13						326

Table 34: Coverage of partial verb chains of NODES projected over French wordnet structure

LEN	DE	EE	$\cap(DE,EE)$	FR
1	3463	3276	3159	3548
2	2379	1780	1416	3380
3	1395	589	437	2569
4	736	153	102	1584
5	299	35	11	809
6	124	4		393
7	49	2		193
8	11			69
9	3			18
10	2			5

Table 35: Coverage of partial verb chains of EDGES projected over French wordnet structure

LEN	DE	EE	$\cap(DE,EE)$	FR
1	1157	1261	343	3380
2	187	208	12	2569
3	47	26		1584
4		2		809

Table 36: Coverage of partial noun chains of NODES projected over Estonian wordnet structure

LEN	DE	FR	CZ	$\cap(DE,FR)$	$\cap(DE,CZ)$	$\cap(FR,CZ)$	$\cap(DE,FR,CZ)$	EE
1	4489	4541	4506	4485	4301	4505	4293	4542
2	3850	4487	3542	3838	2651	3472	2635	4538
3	2708	4329	2219	2658	1314	2147	1298	4475
4	1916	3989	978	1799	412	899	381	4219
5	1185	3222	410	1159	130	336	129	3771
6	588	2283	120	555	34	104	34	2953
7	222	1531	27	209	7	25	7	1989
8	89	875	2	83		1		1333
9	23	442		20				760
10	10	174		10				327
11	5	40		5				92
12		1						10

Table 37: Coverage of partial noun chains of EDGES projected over Estonian wordnet structure

LEN	DE	FR	CZ	$\cap(DE,FR)$	$\cap(DE,CZ)$	$\cap(FR,CZ)$	$\cap(DE,FR,CZ)$	EE
1	2576	4191	1040	2411	166	716	132	4538
2	890	3316	93	750	27	74	25	4475
3	109	2478	19	72	3	14	3	4219
4	25	1761		17				3771
5	1	969		1				2953
6		577						1989
7		246						1333
8		114						760
9		26						327
10		1						92

Table 38: Coverage of partial VERB chains of NODES projected over Estonian wordnet structure

LEN	DE	FR	$\cap(DE,FR)$	EE
1	2060	2089	2038	2109
2	1681	1837	1612	2096
3	1096	1319	1027	1662
4	617	818	562	1076
5	300	437	279	621
6	131	227	120	339
7	49	96	45	158
8	21	37	17	86
9	5	11	5	29
10	1	4	1	14

Table 39 Coverage of partial VERB chains of EDGES projected over Estonian wordnet structure

LEN	DE	FR	$\cap(DE,FR)$	EE
1	635	809	222	2096
2	37	125	9	1662
3	1	13		1076
4		1		621

Table 40: Coverage of partial noun chains of NODES projected over Czech wordnet structure

LEN	DE	FR	EE	$\cap(DE,FR)$	$\cap(DE,EE)$	$\cap(FR,EE)$	$\cap(DE,FR,EE)$	CZ
1	5819	7102	5908	5714	4942	5867	4914	7170
2	3238	6216	3279	3002	1971	3147	1919	7170
3	1183	3806	1212	1004	546	1110	506	5048
4	306	1727	298	243	126	261	112	2396
5	45	599	28	35	7	17	7	890
6	5	141		1				222
7		20						36

Table 41: Coverage of partial noun chains of EDGES projected over Czech wordnet structure

LEN	DE	FR	EE	$\cap(DE,FR)$	$\cap(DE,EE)$	$\cap(FR,EE)$	$\cap(DE,FR,EE)$	CZ
1	732	3472	1225	457	207	931	182	7170
2	70	851	153	40	31	129	29	5048
3	9	298	17	8	2	11	2	2396
4		74						890
5		6						222

### Appendix III Partial chains with 1 gap projected over the German, French, Estonian and Czech wordnet

Table 42: Coverage of partial noun chains of NODES with 1 gap projected over German wordnet structure

LEN	FR	EE	CZ	$\cap(FR,EE)$	$\cap(FR,CZ)$	$\cap(EE,CZ)$	$\cap(FR,EE,CZ)$	DE
3	7411	5965	6109	5855	5835	4954	4877	14457
4	7061	5247	4182	5105	3975	3116	3068	13659
5	5693	3406	1448	3124	1259	510	466	12057
6	2434	876	479	759	424	65	61	9566
7	1353	319	92	290	78	1	1	6940
8	597	69	30	66	26			4843
9	200	16	3	13	3			3320
10	49	1		1				2079
11	3							1230

Table 43: Coverage of partial noun chains of EDGES with 1 gap projected over German wordnet structure

LEN	FR	EE	CZ	$\cap(FR,EE)$	$\cap(FR,CZ)$	$\cap(EE,CZ)$	$\cap(FR,EE,CZ)$	DE
3	1779	1617	30	1455	2	0	0	13659
4	173	94	8	28				12057
5	29	13						9566
6	2							6940

Table 44: Coverage of partial VERB chains of NODES with 1 gap projected over German wordnet structure

LEN	FR	EE	$\cap(FR,EE)$	DE
3	1649	1471	1353	7917
4	1415	943	882	6860
5	901	495	496	5056
6	590	172	156	3853
7	339	112	92	2922
8	132	51	43	1991
9	30	14	13	487
10	4			101
11	1			8

Table 45: Coverage of partial VERB chains of EDGES with 1 gap projected over German wordnet structure

LEN	FR	EE	$\cap(FR,EE)$	DE
3	113	2058	1	6860
4	33	52		5056
5	11			3853

Table 46: Coverage of partial noun chains of NODES with 1 gap projected over French wordnet structure

LEN	DE	EE	CZ	$\cap(DE,EE)$	$\cap(DE,CZ)$	$\cap(EE,CZ)$	$\cap(DE,EE,CZ)$	FR
3	7725	2183	6278	7071	4354	4667	3308	13370
4	6594	2139	4837	5793	2727	3105	1845	13121
5	5006	1988	2966	3427	1049	1537	774	12213
6	3226	1623	1260	2026	285	537	184	10300
7	1665	878	457	1051	96	142	75	7631
8	720	510	135	410	6	62	1	4530
9	260	221	14	136	1			2197
10	63	93		31				1026
11	10	12		5				326
12	4	7		3				141
13	2	3						59

Table 47: Coverage of partial noun chains of EDGES with 1 gap projected over French wordnet structure

LEN	DE	EE	CZ	$\cap(DE,EE)$	$\cap(DE,CZ)$	$\cap(EE,CZ)$	$\cap(DE,EE,CZ)$	FR
3	927	915	130	708	0	8	0	13121
4	502	785	50	426				12213
5	272	600	18	243				10300
6	50	409	2	18				7631
7	7	131		3				4530
8		32						2197
9		13						1026
10		5						326
11		1						141

Table 48: Coverage of partial verb chains of NODES with 1 gap projected over French wordnet structure

LEN	DE	EE	$\cap(DE,EE)$	FR
3	324	793	678	2569
4	202	501	400	1584
5	99	205	166	809
6	56	60	30	393
7	26	13	3	193
8	13	1	1	69
9	2			18

Table 49: Coverage of partial verb chains of EDGES with 1 gap projected over French wordnet structure

LEN	DE	EE	$\cap(DE,EE)$	FR
3	100	18	2	1584
4	17	2		809
5	3			393

Table 50: Coverage of partial noun chains of NODES with 1 gap projected over Estonian wordnet structure

LEN	DE	FR	CZ	$\cap(DE,FR)$	$\cap(DE,CZ)$	$\cap(FR,CZ)$	$\cap(DE,FR,CZ)$	EE
3	2435	361	1849	2387	1329	1791	1282	4475
4	2132	351	1385	2082	768	1318	740	4219
5	1583	318	917	1514	457	864	441	3771
6	1139	294	464	1023	226	421	222	2953
7	754	221	176	719	88	141	84	1989
8	351	146	59	326	16	48	15	1333
9	102	104	11	87		8		760
10	34	44		31				327
11	7	10		7				92
12		3						10

Table 51: Coverage of partial noun chains of EDGES with 1 gap projected over Estonian wordnet structure

LEN	DE	FR	CZ	$\cap(DE,FR)$	$\cap(DE,CZ)$	$\cap(FR,CZ)$	$\cap(DE,FR,CZ)$	EE
3	580	1105	28	535	0	19	0	4219
4	440	892	8	412		8		3771
5	138	629	3	131		3		2953
6	29	402		15				1989
7	3	186		3				1333
8		107						760
9		31						327
10		13						92
11		1						10



Table 52: Coverage of partial VERB chains of NODES with 1 gap projected over Estonian wordnet structure

LEN	DE	FR	$\cap(DE,FR)$	EE
3	220	114	207	1662
4	171	87	160	1076
5	107	62	92	621
6	62	46	51	339
7	29	30	22	158
8	16	25	11	86
9	4	8	3	29
10	1	5		14
11		1		2

Table 53: Coverage of partial VERB chains of EDGES with 1 gap projected over Estonian wordnet structure

LEN	DE	FR	$\cap(DE,FR)$	EE
3	25	61	1	1076
4	2	10		621
5		3		339
6		1		158

Table 54: Coverage of partial noun chains of NODES with 1 gap projected over Czech wordnet structure

LEN	DE	FR	EE	$\cap(DE,FR)$	$\cap(DE,EE)$	$\cap(FR,EE)$	$\cap(DE,FR,EE)$	CZ
3	403	0	145	373	233	127	220	5048
4	174		70	150	56	63	51	2396
5	33		17	28	9	14	8	890
6	6		3	3	1	1		222

Table 55: Coverage of partial noun chains of EDGES with 1 gap projected over Czech wordnet structure

LEN	DE	FR	EE	$\cap(DE,FR)$	$\cap(DE,EE)$	$\cap(FR,EE)$	$\cap(DE,FR,EE)$	CZ
3	3	74	26	0	0	11	0	2396
4		11	1					890
5		2						222