Comparison of the Final Wordnets
Dutch, Spanish and Italian

Version 2, Final
July 16, 1999

Contributors:

Piek Vossen, University of Amsterdam
Salvador Climent, Maria Antonia Marti, Mariona Taule, Universitat de Barcelona
Julio Gonzalo, Irina Chugur, M. Felisa Verdejo, UNED
Gerard Escudero, German Rigau, Horacio Rodriguez, Universitat Politecnica de Catalunya
Antonietta Alonge, Francesca Bertagna, Rita Marinelli, Adriana Roventini, Luca Tarasi, Istituto di Linguistica del CNR, Pisa

Deliverable D029, D030, WP3, WP4
EuroWordNet, LE2-4003

Identification number	LE-4003-D029-D030
Type	Document and Lingware
Title	Comparison of the Final Wordnets Dutch, Spanish and Italian
Status	Draft
Deliverable	D-029, D-030
Work Package	WP3 and WP4
Task	T4
Period covered	January 1999 - March 1999
Date	July 16, 1999
Version	2
Number of pages	34
Authors	⇨ Piek Vossen, Laura Bloksma, University of Amsterdam ⇨ Salvador Climent, Maria Antonia Marti, Mariona Taule, Universitat de Barcelona ⇨ Julio Gonzalo, Irina Chugur, M. Felisa Verdejo, UNED ⇨ Gerard Escudero, German Rigau, Horacio Rodriguez, Universitat Politecnica de Catalunya ⇨ Antonietta Alonge, Francesca Bertagna, Rita Marinelli, Adriana Roventini, Luca Tarasi, Istituto di Linguistica del CNR, Pisa
WP/Task responsible	PSA/FUE
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: Piek.Vossen@hum.uva.nl
EC project officer	Ray Hudson
Status	Public
Actual distribution	Project Consortium, the EuroWordNet User Group, the world via http://www.hum.uva.nl/~ewn.

Suplementary notes	n.a.
Key words	Linguistic Resources, Multilingual Wordnets, Language Engineering
Abstract	
Status of the abstract	Final
Received on	
Recipient's catalogue number	

Executive Summary

This deliverable describes the comparison of the final wordnets for Dutch, Spanish and Italian. The wordnets contain between 25-44K synsets and 50-70K word meanings. This is between 30-50% 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 global coverage and matching of the data.

Table of Contents

1. Introduction 6
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 18
   4.2. Comparison of the ILI-graphs with WordNet1.5 21
5. Conclusions 25
References 26
Appendix I Projection of complete chains on the Dutch, Italian and Spanish wordnets 27
Appendix II Projection of partial chains on the Dutch, Italian and Spanish wordnets 28

List of Tables

Table 1: Intersection of ILI-references in English (WN), Dutch (NL), Italian (IT) and Spanish (ES) 8
Table 2: Overview of Composite ILI-records in the ILI 8
Table 3: The intersection of the English (WN), Dutch (NL), Italian (IT), and Spanish (ES) ILI-references expanded with 9
Composite ILI-records. 9
Table 4: Nominal Synsets clustered as 1stOrder Concepts 11
Table 5: Nominal Synsets clustered as 2ndOrder Concepts 12
Table 6: Verbal Synsets clustered as 2ndOrder Concepts 13
Table 7: Nominal Synsets clustered as 3rdOrder Concepts 13
Table 8: Dutch, Spanish and Italian Nouns clustered over the WordNet1.5 Lexicographer's file codes 14
Table 9: Dutch, Spanish and Italian Verbs clustered over the WordNet1.5 Lexicographer's file codes 15
Table 10: ILI chains for nouns 18
Table 11: ILI chains for verbs 18
Table 12: Frequencies and ratios of noun chains / length /language 20
Table 13: Frequencies and ratios of verb chains / length /language 20
Table 14: Coverage of complete noun chains projected over WN1.5 structure 21
Table 15 Coverage of complete verb chains projected over WN1.5 structure 21
Table 16: Coverage of partial noun chains of NODES projected over WN1.5 structure 22
Table 17: Coverage of partial noun chains of EDGES projected over WN1.5 structure 22
Table 18: Coverage of partial VERB chains of NODES projected over WN1.5 structure 22
Table 19: Coverage of partial VERB chains of EDGES projected over WN1.5 structure 23
Table 20: Comparison in partial coverage of WN1.5 chains by the intersection of WNs between subset1, subset2 and the Final Set. 23
Table 21: Coverage of partial noun chains of NODES with 1 gap projected over WN1.5 structure 24
Table 22: Coverage of partial NOUN chains of EDGES with 1 gap projected over WN1.5 structure 24
Table 23: Coverage of partial VERB chains of NODES with 1 gap projected over WN1.5 structure 24
Table 24: Coverage of partial VERB chains of EDGES with 1 gap projected over WN1.5 structure 24
Table 25: Comparison in 1-gap coverage with WN1.5 chains by the intersection of WNs between subset1, subset2 and the Final Set. 25
Table 26 Coverage of complete noun chains projected over Spanish wordnet structure 27
Table 27 Coverage of complete verb chains projected over Spanish wordnet structure 27
Table 28: Coverage of complete noun chains projected over Dutch wordnet structure 27
Table 29: Coverage of complete verb chains projected over Dutch wordnet structure 27
Table 30: Coverage of complete noun chains projected over Italian wordnet structure 27
Table 31: Coverage of complete verb chains projected over Italian wordnet structure 27
Table 32: Coverage of partial noun chains of NODES projected over Spanish wordnet structure 28
Table 33: Coverage of partial noun chains of EDGES projected over Spanish wordnet structure 28
Table 34: Coverage of partial VERB chains of NODES projected over Spanish wordnet structure 28
Table 35: Coverage of partial VERB chains of EDGES projected over Spanish wordnet structure 28
Table 36: Coverage of partial noun chains of NODES projected over Dutch wordnet structure 28
Table 37: Coverage of partial noun chains of EDGES projected over Dutch wordnet structure 29
Table 38: Coverage of partial verb chains of NODES projected over Dutch wordnet structure 29
Table 39: Coverage of partial verb chains of EDGES projected over Dutch wordnet structure 29
Table 40: Coverage of partial noun chains of NODES projected over Italian wordnet structure 29
Table 41: Coverage of partial noun chains of EDGES projected over Italian wordnet structure 29
Table 42: Coverage of partial VERB chains of NODES projected over Italian wordnet structure 29
Table 43: Coverage of partial VERB chains of EDGES projected over Italian wordnet structure 30
Table 44: Coverage of partial noun chains of NODES with 1 gap projected over Spanish wordnet structure 32
Table 45: Coverage of partial noun chains of EDGES with 1 gap projected over Spanish wordnet structure 32
Table 46: Coverage of partial VERB chains of NODES with 1 gap projected over Spanish wordnet structure 32
Table 47: Coverage of partial VERB chains of EDGES with 1 gap projected over Spanish wordnet structure 32
Table 48: Coverage of partial noun chains of NODES with 1 gap projected over Dutch wordnet structure 32
Table 49: Coverage of partial noun chains of EDGES with 1 gap projected over Dutch wordnet structure 32
Table 50: Coverage of partial verb chains of NODES with 1 gap projected over Dutch wordnet structure 33
Table 51: Coverage of partial verb chains of EDGES with 1 gap projected over Dutch wordnet structure 33
Table 52: Coverage of partial noun chains of NODES with 1 gap projected over Italian wordnet structure 33
Table 53: Coverage of partial noun chains of EDGES with 1 gap projected over Italian wordnet structure 33
Table 54: Coverage of partial VERB chains of NODES with 1 gap projected over Italian wordnet structure 33
Table 55: Coverage of partial VERB chains of EDGES with 1 gap projected over Italian wordnet structure 33

1. Introduction

This deliverable describes the comparison of the final wordnets for Dutch, Spanish and Italian. 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 25K and 45K synsets (see D032D033 for an overview). 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, e.g. in the case of the Dutch wordnet 16% of the nouns and 11% of the verbs have no equivalence link. 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. For example in Dutch, there is an equivalent for the verb "to contain" (which is "bevatten") but not for noun "container", but a mapping can be expressed to the noun from the Dutch verb with an EQ_INVOLVED relation:

"bevatten"
    EQ_SYNONYM "to contain"
    EQ_INVOLVED "container"

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, a more global matching is 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 Spanish, Italian and Dutch wordnet only (which is a bit more than 50% of WN15):

Table 1: Intersection of ILI-references in English (WN), Dutch (NL), Italian (IT) and Spanish (ES)

	Nouns			***Verbs***		
	Total	**62780**	**32520**	**Total**	**12215**	**7455**
	frequency	**% of ∪ (WN,IT,NL,ES)**	**% of ∪ (IT,NL,ES)**	**frequency**	**% of ∪ (WN,IT, NL, ES)**	**% of ∪ (IT, NL, ES)**
ES	24153	38.5%	74.3%	4074	33.4%	54.6%
IT	13950	22.2%	42.9%	3569	29.2%	47.9%
NL	20877	33.3%	64.2%	5562	45.5%	74.6%
∩ (ES, IT)	10449	16.6%	32.1%	2030	16.6%	27.2%
∩ (ES, NL)	14302	22.8%	44.0%	2778	22.7%	37.3%
∩ (IT, NL)	9445	15.0%	29.0%	2574	21.1%	34.5%
∩ (ES, IT, NL)	7736	12.3%	23.8%	1632	13.4%	21.9%

The intersection for nouns between wordnet pairs ranges between 30% and 44% of the total union of ILI-records occurring in all 3 wordnets. Including WordNet1.5, the intersection goes down to 15% up to 23%. This lower coverage is obvious because the total union of the 3 languages is about 50% of WordNet1.5. In the case of verbs, we get similar results: 27% up to 37% intersection between wordnet pairs, compared to the union of 3 languages, and 16% to 23% if we also include WordNet1.5 (maximum coverage is 50%). The intersection of 3 languages is lower, but close to the lowest intersection between language pairs: 24% for nouns and 22% for verbs (out of the union of 3 languages). This corresponds with a set of 7,736 nominal and 1,632 verbal concepts that are (somehow) lexicalized in 4 languages. This intersection includes the set of 1,300 Base Concepts, which is used as a common starting point by all the partners. The union of concepts lexicalized in 3 languages is 18,724 nouns and 4,118 verbs.

As discussed in previous deliverables (D014D015) and papers (Peters et al. 1998, Peters & Peters 1999), the ILI has been adapted to provide a more efficient mapping across languages. Currently, so-called Composite ILI-records have been added that group senses in Wordnet1.5 between which there is a metonymy relation (e.g. *university* as a *building* and an *institute*) or that can be generalized to single more abstract senses (e.g. *fruit* as a *plant-organ* and as *food*). This reduces the sense-differentiation in WordNet1.5. All senses in the local wordnets with a reference to a WordNet1.5 synset that is involved in such a cluster, have automatically received an additional equivalence relation (EQ_METONYM or EQ_GENERALIZATION) to the new ILI-cluster. This means that synsets across wordnets that are linked to different senses of the same cluster can still be mapped via the EQ_METONYM or EQ_GENERALIZATION relations.

Table 2 lists the number of clusters that have been added and the number of words and word senses that are involved.

Table 2: Overview of Composite ILI-records in the ILI

	Metonymy			***Generalization***		
	clusters	words	word senses	words	words	word senses
Nouns	30	24	67	1703	1398	3205
Verbs				2905	1799	5134

Table 3 then shows the effect of expanding the lists of associated ILI-records with the clusters, in each case that at least one sense of the cluster was included. For the nouns we see only a very small increase of about 1 to 1.5%. For example, the total intersection for all 4 languages increased from 7736 (23,8%) to 8183 (25,2%). This is explained by the fact that the clusters only make up a small proportion of the total set of nouns. However, if we look at the verbs we see a doubling of the total intersection: from 1632 (21,9%) to 3051 (40,9%).

Table 3: The intersection of the English (WN), Dutch (NL), Italian (IT), and Spanish (ES) ILI-references expanded with Composite ILI-records.

	Nouns			***Verbs***		
	Total	**62780**	**32520**	**Total**	**12215**	**7455**
	frequency	**% of ∪ (WN,IT, NL, ES)**	**% of ∪ (IT, NL, ES)**	**frequency**	**% of ∪ (WN,IT, NL, ES)**	**% of ∪ (IT, NL, ES)**
ES	24596	39.2%	75.6%	4654	38.1%	62.4%
IT	14272	22.7%	43.9%	4673	38.3%	62.7%
NL	21259	33.9%	65.4%	6416	52.5%	86.1%
∩ (ES, IT)	10907	17.4%	33.5%	3272	26.8%	43.9%
∩ (ES, NL)	14773	23.5%	45.4%	3870	31.7%	51.9%
∩ (IT, NL)	9862	15.7%	30.3%	3950	32.3%	53.0%
∩ (ES, IT, NL)	8183	13.0%	25.2%	3051	25.0%	40.9%

Relatively many more verbal clusters have been added than nominal clusters. About 29% of the verbal senses is involved in a cluster, compared to only 3% of the nominal senses, which explains the stronger effect for verbs. Since the polysemy-rate for verbs is also higher (1.75 senses per verb, 1.21 senses per nouns), there is not much more to gain for the nouns. We can therefore expect a much bigger effect of the verbal clusters in Word-Sense-Disambiguation and Information-Retrieval tasks than for the nouns.

The above figures give the maximal matching across the 4 languages, where it should be noted that some of the ILI-references may be based on complex equivalence relations to local synsets (such as eq_hyperonym, eq_meronym, eq_role, etc.). For cross-language retrieval this may not be a problem. Furthermore, the matching across language-pairs is higher: 30-45% for nouns and 43-53% for verbs.

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 Spanish, Dutch and Italian first subset 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 therefore still gives a good indication of the conceptual coverage.

Top[0]	
1stOrderEntity[1]	**2ndOrderEntity**[0]
Origin[0] Natural[21] Living[30] Plant[18] Human[106] Creature[2] Anima[l23] Artifact[144] **Form**[0] Substance[32] Solid[63] Liquid[13] Gas[1] Object1[62] **Composition**[0] Part[86] Group[63] **Function**[55] Vehicle[8] Representation[12] MoneyRepresentation[10] LanguageRepresentation[34] ImageRepresentation[9] Software[4] Place[45] Occupation[23] Instrument[18] Garment[3] Furniture[6] Covering[8] Container[12] Comestible[32] Building[13]	**SituationType**[6] Dynamic[134] BoundedEvent[183] UnboundedEvent[48] Static[28] Property[61] Relation[38] **SituationComponent**[0] Cause[67] Agentive[170] Phenomenal[17] Stimulating[25] Communication[50] Condition[62] Existence[27] Experience[43] Location[76] Manner[21] Mental[90] Modal[10] Physical[140] Possession[23] Purpose[137] Quantity[39] Social[102] Time[24] Usage[8]
3rdOrderEntity[33]	

Figure 1: The EuroWordNet Top-Ontology

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

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 4, 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 NL, ES and IT gives the percentage of the corresponding TC clusters in WordNet1.5.

Table 4: Nominal Synsets clustered as 1stOrder Concepts

	WN		*NL*			*ES*			*IT*		
Top-Concept	**TC-Tokens**	**%of wn**	**TC-Tokens**	**% of nl**	**%of wn**	**TC-Tokens**	**%of es**	**%of wn**	**TC-Tokens**	**%of it**	**%of wn**
Animal	14068	3.99%	1193	0.97%	8.5%	2458	1.81%	17.5%	1122	1.44%	8.0%
Artifact	19562	5.55%	10803	8.83%	55.2%	9969	7.36%	51.0%	6494	8.34%	33.2%
Building	1022	0.29%	707	0.58%	69.2%	628	0.46%	61.4%	434	0.56%	42.5%
Comestible	3377	0.96%	1393	1.14%	41.2%	1614	1.19%	47.8%	624	0.80%	18.5%
Container	1725	0.49%	778	0.64%	45.1%	799	0.59%	46.3%	432	0.55%	25.0%
Covering	2030	0.58%	1208	0.99%	59.5%	1027	0.76%	50.6%	690	0.89%	34.0%
Creature	664	0.19%	159	0.13%	23.9%	254	0.19%	38.3%	27	0.03%	4.1%
Function	34081	9.68%	17668	14.44%	51.8%	18904	13.96%	55.5%	11043	14.18%	32.4%
Furniture	298	0.08%	171	0.14%	57.4%	147	0.11%	49.3%	87	0.11%	29.2%
Garment	756	0.21%	494	0.40%	65.3%	426	0.31%	56.3%	292	0.37%	38.6%
Gas	93	0.03%	67	0.05%	72.0%	62	0.05%	66.7%	49	0.06%	52.7%
Group	27805	7.90%	3357	2.74%	12.1%	3630	2.68%	13.1%	2337	3.00%	8.4%
Human	11543	3.28%	6372	5.21%	55.2%	7683	5.67%	66.6%	4488	5.76%	38.9%
ImageRepresentation	780	0.22%	412	0.34%	52.8%	426	0.31%	54.6%	294	0.38%	37.7%
Instrument	7036	2.00%	4102	3.35%	58.3%	3590	2.65%	51.0%	2564	3.29%	36.4%
LanguageRepresent.	2844	0.81%	1273	1.04%	44.8%	1218	0.90%	42.8%	691	0.89%	24.3%
Liquid	1629	0.46%	617	0.50%	37.9%	500	0.37%	30.7%	339	0.44%	20.8%
Living	47104	13.37%	10225	8.36%	21.7%	13661	10.08%	29.0%	7408	9.51%	15.7%
MoneyRepresentation	372	0.11%	190	0.16%	51.1%	183	0.14%	49.2%	111	0.14%	29.8%
Natural	68370	19.41%	21948	17.94%	32.1%	24556	18.13%	35.9%	14400	18.49%	21.1%
Object	48162	13.68%	20206	16.51%	42.0%	22608	16.69%	46.9%	13242	17.00%	27.5%
Occupation	2059	0.58%	1209	0.99%	58.7%	1395	1.03%	67.8%	824	1.06%	40.0%
Part	12083	3.43%	4806	3.93%	39.8%	5819	4.30%	48.2%	2586	3.32%	21.4%
Place	5281	1.50%	2072	1.69%	39.2%	2439	1.80%	46.2%	1227	1.58%	23.2%
Plant	18874	5.36%	1534	1.25%	8.1%	2012	1.49%	10.7%	1121	1.44%	5.9%
Representation	934	0.27%	560	0.46%	60.0%	577	0.43%	61.8%	302	0.39%	32.3%
Software	201	0.06%	80	0.07%	39.8%	91	0.07%	45.3%	49	0.06%	24.4%
Solid	6319	1.79%	2845	2.33%	45.0%	2721	2.01%	43.1%	1406	1.81%	22.3%
Substance	12365	3.51%	5447	4.45%	44.1%	5599	4.13%	45.3%	2847	3.66%	23.0%
Vehicle	747	0.21%	466	0.38%	62.4%	466	0.34%	62.4%	352	0.45%	47.1%
Total	352184		122362		34.7%	135462		38.5%	77882		22.1%

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 are 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 33%, since the aimed total size of the wordnets is about 1/3 of WordNet1.5. However, 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. 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.

First of all we see, as expected, that Creature, Animal, and Plant are less well covered in all 3 wordnets, if compared to WordNet1.5. Another, unexpected, case of under-representation is Group. For Spanish and Dutch, all other clusters are well-represented, even above the 33% on average. In general we can say that the Dutch and Spanish wordnets are well-balanced with respect to WordNet1.5 and also with respect to each other. Some classes are even over-represented: Building, Gas, and Occupation. The Italian clustering is slightly lower (22% average) but the clustering is reasonably balanced, except for Comestible (18.5%!). The lower coverage is probably caused by a lack of equivalence relations because the size of the Italian wordnet is compatible with the others.

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

Table 5: Nominal Synsets clustered as 2ndOrder Concepts

	WN		*NL*			*ES*			*IT*		
Top-Concept	**TC-Tokens**	**%of wn**	**TC-Tokens**	**% of nl**	**%of wn**	**TC-Tokens**	**%of es**	**%of wn**	**TC-Tokens**	**%of it**	**%of wn**
Agentive	12255	6.84%	6311	7.10%	51.5%	7408	7.07%	60.4%	5408	8.09%	44.1%
BoundedEvent	8142	4.55%	4245	4.77%	52.1%	4856	4.64%	59.6%	3523	5.27%	43.3%
Cause	15458	8.63%	8046	9.05%	52.1%	9305	8.89%	60.2%	6576	9.84%	42.5%
Communication	7097	3.96%	3624	4.08%	51.1%	3981	3.80%	56.1%	2365	3.54%	33.3%
Condition	3951	2.21%	2003	2.25%	50.7%	2342	2.24%	59.3%	1620	2.42%	41.0%
Dynamic	20026	11.18%	10519	11.83%	52.5%	12226	11.68%	61.1%	8301	12.42%	41.5%
Existence	330	0.18%	242	0.27%	73.3%	202	0.19%	61.2%	187	0.28%	56.7%
Experience	6862	3.83%	3558	4.00%	51.9%	4268	4.08%	62.2%	2540	3.80%	37.0%
Location	1536	0.96%	868	0.88%	56.5%	788	0.75%	51.3%	746	1.23%	48.6%
Manner	934	0.52%	469	0.53%	50.2%	567	0.54%	60.7%	369	0.55%	39.5%
Mental	10444	5.83%	5212	5.86%	49.9%	6158	5.88%	59.0%	3681	5.51%	35.2%
Modal	542	0.30%	278	0.31%	51.3%	352	0.34%	64.9%	233	0.35%	43.0%
Phenomenal	2132	1.19%	1099	1.24%	51.5%	1204	1.15%	56.5%	683	1.02%	32.0%
Physical	8066	4.50%	4168	4.69%	51.7%	4541	4.34%	56.3%	2964	4.43%	36.7%
Possession	1411	0.79%	714	0.80%	50.6%	647	0.62%	45.9%	418	0.63%	29.6%
Property	12336	6.89%	5542	6.23%	44.9%	7777	7.43%	63.0%	4928	7.37%	39.9%
Purpose	15275	8.53%	7435	8.36%	48.7%	8340	7.96%	54.6%	5321	7.96%	34.8%
Quantity	3864	2.16%	1649	1.85%	42.7%	1977	1.89%	51.2%	900	1.35%	23.3%
Relation	6822	3.81%	3235	3.64%	47.4%	3677	3.51%	53.9%	2132	3.19%	31.3%
Social	12024	6.71%	5765	6.48%	47.9%	6610	6.31%	55.0%	3840	5.74%	31.9%
Static	21365	11.93%	9777	11.00%	45.8%	12506	11.94%	58.5%	7623	11.40%	35.7%
Stimulating	1119	0.62%	588	0.66%	52.5%	721	0.69%	64.4%	433	0.65%	38.7%
Time	1444	0.81%	720	0.81%	49.9%	871	0.83%	60.3%	266	0.40%	18.4%
UnboundedEvent	4567	2.55%	2472	2.78%	54.1%	2981	2.85%	65.3%	1726	2.58%	37.8%
Usage	1084	0.61%	364	0.41%	33.6%	406	0.39%	37.5%	68	0.10%	6.3%
Total	179086		88903		49.6%	104711		58.5%	66851		37.3%

Table 6: Verbal Synsets clustered as 2ndOrder Concepts

	WN		NL			ES			IT		
Top-Concept	**TC-Tokens**	**%of wn**	**TC-Tokens**	**% of nl**	**%of wn**	**TC-Tokens**	**%of es**	**%of wn**	**TC-Tokens**	**%of it**	**%of wn**
Agentive	8176	7.1%	4947	7.2%	60.5%	3248	6.5%	39.7%	3415	7.2%	41.8%
BoundedEvent	10262	8.9%	6145	8.9%	59.9%	4410	8.8%	43.0%	4391	9.3%	42.8%
Cause	15261	13.2%	9136	13.3%	59.9%	6468	12.9%	42.4%	6446	13.6%	42.2%
Communication	3969	3.4%	2425	3.5%	61.1%	1666	3.3%	42.0%	1825	3.9%	46.0%
Condition	1730	1.5%	1093	1.6%	63.2%	828	1.7%	47.9%	820	1.7%	47.4%
Dynamic	23487	20.4%	13987	20.3%	59.6%	10182	20.3%	43.4%	9532	20.2%	40.6%
Existence	2296	2.0%	1349	2.0%	58.8%	912	1.8%	39.7%	1145	2.4%	49.9%
Experience	2067	1.8%	1312	1.9%	63.5%	1199	2.4%	58.0%	927	2.0%	44.8%
Location	8184	7.1%	4778	6.9%	58.4%	3757	7.5%	45.9%	2799	5.9%	34.2%
Manner	350	0.3%	192	0.3%	54.9%	139	0.3%	39.7%	89	0.2%	25.4%
Mental	3048	2.6%	1840	2.7%	60.4%	1528	3.0%	50.1%	1370	2.9%	44.9%
Modal	101	0.1%	58	0.1%	57.4%	72	0.1%	71.3%	47	0.1%	46.5%
Phenomenal	129	0.1%	108	0.2%	83.7%	86	0.2%	66.7%	55	0.1%	42.6%
Physical	11642	10.1%	6985	10.1%	60.0%	5408	10.8%	46.5%	4517	9.6%	38.8%
Possession	1968	1.7%	1195	1.7%	60.7%	922	1.8%	46.8%	847	1.8%	43.0%
Property	504	0.4%	294	0.4%	58.3%	294	0.6%	58.3%	227	0.5%	45.0%
Purpose	4436	3.8%	2670	3.9%	60.2%	1652	3.3%	37.2%	1836	3.9%	41.4%
Quantity	690	0.6%	396	0.6%	57.4%	330	0.7%	47.8%	256	0.5%	37.1%
Relation	960	0.8%	584	0.8%	60.8%	422	0.8%	44.0%	378	0.8%	39.4%
Social	5706	4.9%	3318	4.8%	58.1%	2014	4.0%	35.3%	2371	5.0%	41.6%
Static	6217	5.4%	3434	5.0%	55.2%	2775	5.5%	44.6%	2155	4.6%	34.7%
Stimulating	878	0.8%	548	0.8%	62.4%	590	1.2%	67.2%	403	0.9%	45.9%
Time	98	0.1%	51	0.1%	52.0%	37	0.1%	37.8%	25	0.1%	25.5%
UnboundedEvent	2536	2.2%	1613	2.3%	63.6%	961	1.9%	37.9%	1161	2.5%	45.8%
Usage	646	0.6%	396	0.6%	61.3%	269	0.5%	41.6%	249	0.5%	38.5%
Total	115341		68854		59.7%	50169		43.5%	47286		41.0%

The results are better for the 2ndOrderEntities than for the 1stOrderEntities:

- Dutch and Spanish have an average coverage of 49% and 58% for nouns and 59% and 43% for verbs, respectively, which is much higher than the 33%. The coverage for Italian is 37% for nouns and 41% for verbs;
- No Spanish and Dutch clusters below 33%;
- The Italian wordnet scores low for Quantity, Time, Usage and Manner;
- Spanish and Dutch score extremely high for Existence, Stimulating, Modal and Phenomenal;
- the proportion of the Dutch verbs is relatively high, but this is because there are relatively more verbs in the Dutch wordnet;

Finally, the next table gives the nominal synsets classified as 3rdOrderEntities, where the percentage give the proportion of the set in WordNet1.5. Here we see all 3 wordnets score less than 33%, but Italian scores extremely low with 4%.

Table 7: Nominal Synsets clustered as 3rdOrder Concepts

	WN	AMS		FUE		PSA	
	TC-Tokens	**TC-Tokens**	**% of wn**	**TC-Tokens**	**% of wn**	**TC-Tokens**	**% of wn**
3rdOrderEntity	8059	1388	17.22%	1912	23.73%	340	4.22%

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 8: Dutch, Spanish and Italian Nouns clustered over the WordNet1.5 Lexicographer's file codes

		WN		NL		ES		IT	
	Lexicographer's file code	**TC-Tokens**	**% of wn**	**TC-Tokens**	**% wn**	**TC-Tokens**	**% wn**	**TC-Tokens**	**% wn**
4	**noun.act**	8582	6.83%	4293	50.02%	4912	57.24%	3978	46.35%
5	**noun.animal**	13803	10.99%	1048	7.59%	2311	16.74%	1057	7.66%
6	**noun.artifact**	14994	11.94%	9054	60.38%	8117	54.13%	5550	37.01%
7	**noun.attribute**	4741	3.78%	1722	36.32%	3039	64.10%	1937	40.86%
8	**noun.body**	2900	2.31%	1229	42.38%	1571	54.17%	770	26.55%
9	**noun.cognition**	3997	3.18%	2152	53.84%	2542	63.60%	1466	36.68%
10	**noun.communication**	6819	5.43%	3439	50.43%	3847	56.42%	2268	33.26%
11	**noun.event**	1389	1.11%	748	53.85%	904	65.08%	696	50.11%
12	**noun.feeling**	758	0.60%	345	45.51%	550	72.56%	394	51.98%
13	**noun.food**	3352	2.67%	1368	40.81%	1589	47.40%	601	17.93%
14	**noun.group**	13728	10.93%	1408	10.26%	1310	9.54%	937	6.83%
15	**noun.location**	3231	2.57%	1020	31.57%	1445	44.72%	530	16.40%
16	**noun.motive**	53	0.04%	23	43.40%	33	62.26%	28	52.83%
17	**noun.object**	4083	3.25%	1592	38.99%	2044	50.06%	1016	24.88%
18	**noun.person**	9356	7.45%	5281	56.45%	6776	72.42%	3794	40.55%
19	**noun.phenomenon**	751	0.60%	415	55.26%	355	47.27%	203	27.03%
20	**noun.plant**	18536	14.76%	1367	7.37%	1817	9.80%	1055	5.69%
21	**noun.possession**	1240	0.99%	573	46.21%	541	43.63%	323	26.05%
22	**noun.process**	1038	0.83%	488	47.01%	586	56.45%	353	34.01%
23	**noun.quantity**	2021	1.61%	778	38.50%	890	44.04%	388	19.20%
24	**noun.relation**	944	0.75%	417	44.17%	516	54.66%	237	25.11%
25	**noun.shape**	633	0.50%	312	49.29%	349	55.13%	278	43.92%
26	**noun.state**	3162	2.52%	1495	47.28%	1819	57.53%	1254	39.66%
27	**noun.substance**	4048	3.22%	1938	47.88%	1789	44.19%	962	23.76%
28	**noun.time**	1427	1.14%	705	49.40%	855	59.92%	255	17.87%
	Total	125586		43210	34.41%	50507	40.22%	30330	24.15%

Table 9: Dutch, Spanish and Italian Verbs clustered over the WordNet1.5 Lexicographer's file codes

		WN		NL		ES		IT	
	Lexicographer's file code	TC-Tokens	% of wn	TC-Tokens	% wn	TC-Tokens	% wn	TC-Tokens	% wn
29	verb.body	1095	0.39%	719	65.66%	505	46.12%	484	44.20%
30	verb.change	6379	2.24%	3732	58.50%	2753	43.16%	2571	40.30%
31	verb.cognition	1986	0.70%	1165	58.66%	860	43.30%	875	44.06%
32	verb.communication	3569	1.26%	2238	62.71%	1538	43.09%	1657	46.43%
33	verb.competition	791	0.28%	333	42.10%	176	22.25%	168	21.24%
34	verb.consumption	569	0.20%	366	64.32%	252	44.29%	227	39.89%
35	verb.contact	4028	1.42%	2248	55.81%	1808	44.89%	1272	31.58%
36	verb.creation	1658	0.58%	966	58.26%	675	40.71%	827	49.88%
37	verb.emotion	789	0.28%	488	61.85%	530	67.17%	396	50.19%
38	verb.motion	3865	1.36%	2386	61.73%	1724	44.61%	1348	34.88%
39	verb.perception	870	0.31%	520	59.77%	449	51.61%	323	37.13%
40	verb.possession	1815	0.64%	1134	62.48%	843	46.45%	770	42.42%
41	verb.social	4209	1.48%	2593	61.61%	1458	34.64%	1886	44.81%
42	verb.stative	1345	0.47%	699	51.97%	585	43.49%	494	36.73%
43	verb.weather	117	0.04%	76	64.96%	75	64.10%	50	42.74%
	Total	284257		106083	37.32%	115245	40.54%	74008	26.04%

We see here the same tendencies. For nouns, animal, plant and group are lower, abstract nouns are slightly higher than the expected 33%. Italian scores lower for food, time and location. For the rest, the distribution is reasonably balanced and sufficient. Feeling is relatively high. For verbs, we see that WordNet1.5 scores relatively lower. Hardly any distribution is below 33%, except for Italian competition. High scores for Dutch body and Spanish motion.

4. Comparison of the hyponymy structures

The previous comparison only indicates 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 (NL, IT, SP) 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) ⇐ bewegen (move intransitive) ⇐ bewegen (move reflexive) ⇐ voortbewegen (move location) ⇐ verplaatsen (move from A to B) ⇐ stijgen (move to a higher position) ⇐ 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 as a hyponym the equivalent to ILI record number 01046072 (bewegen) and this one has as hyponym the equivalent to ILI record number 01046072, etc. It should be noted that the ILI-chains are in

many way 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 are 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 (e.g. the nouns in the Spanish wordnet hardly covers 30% of the nouns in WN1.5). 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 other wordnets have unified the verb hierarchy in 2 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 4 times, taking each wordnet as a starting point: WN1.5, Dutch-WN, Italian-WN and Spanish-WN. 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. 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 10: ILI chains for nouns

	ILI nodes	*Tops*	*Leaves*		*Internal Nodes*		*EDGES*	*CHAINS*
WN15	60557	11	47110	77,79%	13436	22,19%	61123	53467
ES	24215	11	18273	75,46%	5931	24,49%	24590	22093
NL	23903	12	19476	81,48%	5663	23,69%	29872	50042
IT	23617	21	21343	90,37%	4427	18,74%	57417	173637

Table 11: ILI chains for verbs

	ILI nodes	*Tops*	*Leaves*		*Internal Nodes*		*EDGES*	*CHAINS*
WN15	11363	573	8446	74,33%	2580	22,71%	10816	8486
ES	4079	366	2927	71,76%	957	23,46%	3728	2948
NL	5865	2	4725	80,56%	1797	30,64%	8655	9965
IT	6478	2	5351	82,60%	1857	28,67%	14827	63631

The number of nodes in 3 wordnets is reasonably equal and covers more than 33% of the nouns and verbs (which is the minimally aimed size). The Italian and Dutch verb nodes are a bit higher, mainly due to the fact that most ILIs are generated by automatic procedures of which the best two matches are selected.

If we look at the number of tops, we see that there are only a few noun-tops in all 4 wordnets, but that only the Dutch and Italian wordnet also have a few verb tops. A limited number of tops is considered to be a good property, since it indicates that the highest levels of the wordnets are somehow classified and the whole structure can be accessed top-down from a few nodes.

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. We can see that the ratios are relatively equal across the wordnets, where there is a tendency for Dutch and Italian to have 10% more leaves for both nouns and verbs but not for the nouns. This can indicate less complexity for noun hierarchies in Italian and Dutch.[1]

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 Spanish wordnets (which is closely related to WordNet1.5) has a similar proportion. If, on the other hand, we look at Italian and Dutch, we see that the number of chains is 2 and 9 times as high. This extreme tangledness of the ILI-chains is due to all the 3 causes. First of all, multiple hyperonyms have been encoded far more systematically encoded, e.g. to deal with Dutch verb compounds such as "dichttrekken" (to close by pulling) which are both linked to "dichtmaken" (close) and "trekken" (pull). Secondly, multiple translations have been chosen when the translations are generated automatically. This both leads to alternative ILI-chains for each translation, but also to the fact that different synsets share the same ILI-records, thus creating more tangled structures.

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. The next example from the Italian wordnet contains 8 edges from which 16 different chains can be constructed:

00016649 00527228
00016649 00528736
00021098 00527228
00021098 00528736
00527228 00542253
00527228 00543162
00528736 00542253
00528736 00543162

We clearly see that this is the case for Dutch and Italian, whereas WordNet1.5 and the Spanish wordnet have less chains than edges.

[1] The fact that the number of internal nodes and leaves exceeds the total number of nodes is due to the fact that some ILI-records can be leaves in one chain and internal nodes in other chains. In that case they are counted twice.

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

Table 12: Frequencies and ratios of noun chains / length /language

	WN		*ES*		*NL*		*IT*	
	frequency	*%*	*frequency*	*%*	*frequency*	*%*	*frequency*	*%*
1					1	0.00		
2	33	0.06	47	0.21	81	0.16	3662	2.11
3	521	0.97	624	2.82	1000	2.00	21344	12.29
4	2220	4.15	1691	7.65	5264	10.52	36975	21.29
5	5664	10.59	3618	16.38	12465	24.91	38892	22.40
6	12730	23.81	4974	22.51	12657	25.29	25622	14.76
7	11741	21.96	4961	22.46	9479	18.94	23870	13.75
8	8737	16.34	3136	14.19	5514	11.02	6845	3.94
9	5940	11.11	1634	7.40	2303	4.60	7873	4.53
10	3305	6.18	889	4.02	916	1.83	6843	3.94
11	1400	2.62	321	1.45	251	0.50	1695	0.98
12	517	0.97	111	0.50	86	0.17	14	0.01
13	364	0.68	68	0.31	23	0.05	2	0.00
14	213	0.40	15	0.07	2	0.00		
15	75	0.14	4	0.02				
16	7	0.01						
Total	53467	100	22093	100	50042	100	173637	100
Average	7.19		6.61		6.13		5.46	

Table 13: Frequencies and ratios of verb chains / length /language

	WN		*ES*		*NL*		*IT*	
	frequency	*%*	*frequency*	*%*	*frequency*	*%*	*frequency*	*%*
1	236	2.78	171	5.80				
2	1867	22.00	798	27.07	15	0.15	163	0.26
3	2530	29.81	883	29.95	218	2.19	1037	1.63
4	1959	23.09	593	20.12	790	7.93	2260	3.55
5	1029	12.13	298	10.11	1895	19.02	8693	13.66
6	462	5.44	125	4.24	2181	21.89	14809	23.27
7	250	2.95	50	1.70	1809	18.15	14971	23.53
8	109	1.28	27	0.92	1316	13.21	10750	16.89
9	32	0.38	1	0.03	927	9.30	6260	9.84
10	10	0.12	2	0.07	508	5.10	3728	5.86
11	2	0.02			206	2.07	960	1.51
12					71	0.71		
13					24	0.24		
14					5	0.05		
Total	8486	100	2948	100	9965	100	63631	100
Average	3.58		3.26		6.68		6.91	

For nouns, we see here that the average length is rather close across the wordnets. Obviously, WordNet1.5 has more depth, but that is what we would have expected since it is 2 up to 3 times the size of the other wordnets. Still, the average depth is 6, also for Spanish and Dutch, and 5 for Italian. Apparently, there is a similar lexicalization expansion at a similar depth. This is in line with predictions made by Rosch (1977) and Berlin (1972) on the need for concepts at the so-called Basic Level.

For verbs, the situation is very different. Here, we see extreme differences between WordNet1.5 and Spanish on the one hand and Dutch and Italian on the other. The depth of the latter twice as high. The main explanation for this is that the top-levels of Dutch and Italian have more structure.

Both Spanish and WordNet1.5 have a number of tops (length 1) that corresponds with level 3 in size of the Dutch wordnet. The Dutch wordnet only has two tops for verbs. It may be that adding this top-classification to WordNet1.5 and Spanish would result in a similar distribution in levels as in Dutch. For the Italian wordnet, the distribution is similar but the absolute frequency is much higher. This is mainly due to the fact that many more chains have been generated for each automatically derived translation.

4.2. Comparison of the ILI-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 14: Coverage of complete noun chains projected over WN1.5 structure

	nodes	*(53467)*	*edges*	*(53467)*
	frequency	*%*	*frequency*	*%*
ES	14221	26.60	14221	26.60
NL	650	1.22	17	0.03
IT	2760	5.16	49	0.09
∩ (ES,NL)	352	0.66	10	0.02
∩ (ES,IT)	1563	2.92	34	0.06
∩ (NL,IT)	190	0.36	0	0.00
∩ (ES,NL,IT)	136	0.25	0	0.00

Table 15 Coverage of complete verb chains projected over WN1.5 structure

	nodes	*(8486)*	*edges*	*(8486)*
	frequency	*%*	*frequency*	*%*
ES	1963	23.13	1963	23.13
NL	1269	14.95	237	2.79
IT	1334	15.72	251	2.96
∩ (ES,NL)	482	5.68	94	1.11
∩ (ES,IT)	553	6.52	123	1.45
∩ (NL,IT)	359	4.23	48	0.57
∩ (ES,NL,IT)	187	2.20	21	0.25

We see that the Spanish wordnet, which is built by expanding WordNet1.5, is very similar to WordNet1.5. Given the fact that the size is 33% of WordNet1.5, the figures 26% and 23% for nouns and verbs respectively, are very high. This also indicates that the coverage and matching is concentrated in the highest regions of the hierarchy. If there is a difference at the top-node, none of the complete chains can have an edge coverage. Consequently, the differences in Spanish are due to the smaller size.

A completely different situation holds for the Dutch and Italian wordnets, which have a hyponymy structure that is totally independent of WordNet1.5. Given the fact that they have only a slightly lower number of nodes than the Spanish wordnet, and given the common approach to build all the wordnets from the same set of Base Concepts top-down, we can only explain the difference by differences at the highest level or by many differences distributed over lower levels. If a few fundamental choices at the top level are different, then it may still be the case that the hyponymy structures are the same at lower levels.

Because of the low overlap of Dutch and Italian with WordNet1.5, it is obvious that the intersection is extremely low. Only 21 verbal chains and 0 nominal chains show full overlap in edges across the four wordnets.

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

LENGTH	ES	NL	IT	∩ (ES,NL)	∩ (ES,IT)	∩ (NL,IT)	∩ (ES,NL,IT)	WN
1	53467	53213	53456	53148	53452	52862	52803	53467
2	53385	43161	47346	41959	47138	40893	40636	53467
3	51541	26862	44076	25162	42764	21573	21089	53434
4	47930	15032	27878	13106	26260	7808	7112	52913
5	42049	6771	21019	5454	19433	2996	2506	50693
6	27582	2781	14817	1929	12552	949	799	45029
7	16789	967	7865	726	6259	169	148	32299
8	8337	196	3526	87	2648	17	12	20558
9	3800	6	1062	3	779			11821
10	1647		380		311			5881
11	647		82		73			2576
12	299		28		25			1176
13	115							659
14	19							295
15	2							82

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

LENGTH	ES	NL	IT	∩ (ES,NL)	∩ (ES,IT)	∩ (NL,IT)	∩ (ES,NL,IT)	WN
1	53385	24777	42395	23412	42167	16076	15994	53467
2	51541	7530	24693	7032	23374	1140	1113	53434
3	47930	582	9081	398	8888	113	113	52913
4	42049	80	1282	43	1245			50693
5	27582	1	83	1	76			45029
6	16789		9		9			32299
7	8337							20558
8	3800							11821
9	1647							5881
10	647							2576
11	299							1176
12	115							659
13	19							295
14	2							82

Table 18: Coverage of partial VERB chains of NODES projected over WN1.5 structure

LENGTH	ES	NL	IT	∩ (ES,NL)	∩ (ES,IT)	∩ (NL,IT)	∩ (ES,NL,IT)	WN
1	8049	7962	7967	7525	7691	7460	7196	8486
2	5973	5020	5434	3968	4712	3581	3126	8250
3	3630	2200	2972	1536	2405	1285	1072	6383
4	1774	767	1417	408	1095	343	237	3853
5	777	174	544	78	379	64	39	1894
6	256	32	192	7	75	5	3	865
7	75	8	37	1	7	1	1	403
8	23	1	3					153
9	2							44
10	1							12

Table 19: Coverage of partial VERB chains of EDGES projected over WN1.5 structure

LENGTH	ES	NL	IT	∩ (ES,NL)	∩ (ES,IT)	∩ (NL,IT)	∩ (ES,NL,IT)	WN
1	5973	2348	2762	1986	2488	719	654	8250
2	3630	330	407	270	345	15	12	6383
3	1774	11	47	4	38			3853
4	777	2	2		1			1894
5	256							865
6	75							403
7	23							153
8	2							44
9	1							12

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. However, if we look at the chains of length 3, for example, we see that there are 51,541 WN1.5 chains of length 3 that can be covered with nodes coming from the Spanish wordnet. For Dutch and Italian, these are 26,862 and 44076 respectively. Dutch determines here the upper limit, and we see that 21,089 nominal chains of length 3 in WordNet1.5 are covered by nodes from all the three languages. For verbs this is 1072 out of an upper limit of 2200 nodes. This means that in principle it would be possible to create WordNet1.5 compatible hierarchies in all the 3 languages with the size of 21,089 and 1,072 nodes and a length of 3. The question is if such a WordNet1.5 hierarchy is also desirable in the different languages.

The other tables with edge coverage than show how compatible the sub-sequences are in terms of the hyponymy relations. Edge coverage for sub-chains is extremely low. For nouns, we see that 15,994 nodes intersect for the 4 languages, but that only 1,113 also share the next hyponymy level, and 113 also a third hyponymy level across 4 languages. For verbs, there are only 654 shared leaves and in 12 cases also a shared 2nd level.

In the next table, we can see how the overlap of partial node chains has increased during the building of the wordnets. Three measurements have been made at 3 points during the project. The core wordnets (10,000 up to 20,000 synsets) built around the Base Concepts represent subset1. Subset 2 is a major extension to the full size (about 30,000 synsets on average) and the final wordnets contain improvements with respect to subset 2.

Table 20: Comparison in partial coverage of WN1.5 chains by the intersection of WNs between subset1, subset2 and the Final Set.

Length	*intersection subset1*	*intersection subset2*	*final intersection*	*first increment*	*second increment*	*% first increment*	*% second increment*
1	*30909*	*51270*	*52803*	*20361*	*1533*	*66*	*2,99*
2	*16151*	*24614*	*40636*	*8463*	*16022*	*52*	*65,09*
3	*6756*	*13568*	*21089*	*6812*	*7521*	*100*	*55,43*
4	*2001*	*8203*	*7112*	*6202*	*-1091*	*310*	*-13,3*
5	*780*	*2826*	*2506*	*2046*	*-320*	*262*	*-11,32*
6	*393*	*1476*	*799*	*1083*	*-677*	*275*	*-45,87*
7	*228*	*462*	*148*	*234*	*-314*	*103*	*-67,97*
8	*9*	*257*	*12*	*248*	*-245*	*275*	*-95,33*
9		*32*		*32*	*-32*		*-100*
10		*2*		*2*	*-2*		*-100*

We clearly see here an increase in intersection as the wordnets grow, but also in the final phase, the

wordnets have increased in overlap with respect to sub-chains of length 2 and 3. However, this gain also results in a decrease of longer chains. The explanation of this is that the final phase has lead to removal of many spurious and wrong translation that have been generated. This resulted in a partial untangling of the hierarchy and thus shorter chains.

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 Dutch, Italian and Spanish WN structure.

Table 21: Coverage of partial noun chains of NODES with 1 gap projected over WN1.5 structure

LENGTH	ES	NL	IT	∩ (ES,NL)	∩ (ES,IT)	∩ (NL,IT)	∩ (ES,NL,IT)	WN
3	7804	29355	12152	28312	11619	20886	20439	53434
4	7776	26152	11616	24655	11086	17228	16775	52913
5	7333	18633	10480	16712	9652	11136	10561	50693
6	6296	12019	7782	10158	6879	6023	5262	45029
7	5017	5326	4602	3866	4119	2531	1960	32299
8	3392	1891	2456	1046	2131	704	560	20558
9	1914	487	1166	268	986	115	98	11821
10	1038	83	538	32	485	11	7	5881
11	564	2	173	1	163			2576
12	232		108		101			1176
13	98		35		4			659
14	43		2					295
15	5							82
16	2							7

Table 22: Coverage of partial NOUN chains of EDGES with 1 gap projected over WN1.5 structure

LENGTH	ES	NL	IT	∩ (ES,NL)	∩ (ES,IT)	∩ (NL,IT)	∩ (ES,NL,IT)	WN
3	0	1180	8927	1011	8636	482	474	52913
4		555	5048	469	4600	199	195	50693
5		130	3683	67	3568	45	45	45029
6		3	1482		1419			32299
7			112		105			20558

Table 23: Coverage of partial VERB chains of NODES with 1 gap projected over WN1.5 structure

LENGTH	ES	NL	IT	∩ (ES,NL)	∩ (ES,IT)	∩ (NL,IT)	∩ (ES,NL,IT)	WN
3	501	1249	950	1001	747	1057	872	6383
4	302	711	533	513	370	460	388	3853
5	188	266	289	176	212	117	88	1894
6	104	81	181	44	131	38	25	865
7	58	15	97	16	62	11	10	403
8	19	1	37	1	11	1	1	153
9	7		8		4			44
10	2		2		2			12
11	1							2

Table 24: Coverage of partial VERB chains of EDGES with 1 gap projected over WN1.5 structure

LENGTH	ES	NL	IT	∩ (ES,NL)	∩ (ES,IT)	∩ (NL,IT)	∩ (ES,NL,IT)	WN
3	0	64	25	32	16	1	0	3853
4		5	8	2	5			1894

The results are better than for complete chains but slightly less then for sub-sequences. In the case of length 4, we are dealing with 16,775 nominal and 388 verbal chains in WordNet1.5 that can be covered with nodes from Dutch, Italian and Spanish, expect for an intermediate node or level that is missing. Although we cannot judge the relevance of this intermediate level, there is thus a potential

for larger overlap across the wordnets.

If these figures are compared with previous measurements, we see that the 1 gap overlap has increased enormously for short sequences.

Table 25: Comparison in 1-gap coverage with WN1.5 chains by the intersection of WNs between subset1, subset2 and the Final Set.

Length	Subset1	Subset2	Final
3	5,672	8,837	20,439
4	4,127	6,444	16,775
5	2,901	5,095	10561
6	227	2,883	5262
7	11	622	1960
8	3	238	560
9	3	65	98

For the shorter chains we see here more than a doubling of the intersection. In the case of verbs, we see a similar difference: from 591 in subset 2 to 872 in the final wordnets for length 3, and from 105 to 388 for length 4.

5. Conclusions

In this document we described the compatibility of the Dutch, Spanish and Italian 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 12% and 13% for nouns and verbs respectively, where the maximal intersection is 33% given the fact that the wordnets are 1/3rd of the size of WordNet1.5. If we look at the union of the ILI-references for the 3 languages, intersection is about 24% for nouns and 22% for verbs. Applying the ILI-clusters to these collections, these percentages increase to 25% and 41% for nouns and verbs respectively.

The figures give the maximal matching across the 4 languages, regardless of the type of equivalence relation. The matching across language-pairs is higher: 30-45% for nouns and 43-53% 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 very 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. In a few fields the Italian wordnet scored lower than average.

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 (33% of WordNet1.5) and the different choices at the top levels of the hierarchy. A much closer match was found for the Spanish wordnet, which is reasonably given the approach to build it by expanding WordNet15 rather than building an independent hierarchy. Looking at sub-chains and chains with one gap we still have measured little overlap in hyponymy relations (edge coverage) across the wordnets. The structural difference between especially the Dutch and Italian wordnets as compared to the Spanish and English wordnet is considerable, or the

translations are extremely unreliable. It is not possible to draw any further conclusions from these figures. Looking at the node coverage, it is clear that potential larger overlap is possible, since a large proportion of hyponymy relations can be covered. Nevertheless, a positive conclusion is that the overlap has increased during the development of the wordnets.

Finally, the relatively large node coverage gives the option to project the WordNet1.5 hyponymy structure on substantial proportions of any of the other wordnets, and vice versa.

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.

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

Miller G.A, R. Beckwidth, 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.

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

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. Rodriquez, C. Peters, A. Roventini, F. Bertagna, A. Alonge, W. Peters.
1998 The Restructured Core wordnets in EuroWordNet: Subset1. EuroWordNet (LE 4003), Deliverable D014D015, University of Amsterdam.

Vossen, P.
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. Rodriquez, A. Alonge, F. Bertagna.
1998 EuroWordNet Restructured Subset2 for Dutch, Spanish and Italian. EuroWordNet (LE 4003), Deliverable D027D028, University of Amsterdam.

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

Appendix I Projection of complete chains on the Dutch, Italian and Spanish wordnets

Table 26 Coverage of complete noun chains projected over Spanish wordnet structure

	nodes	*(22093)*	*edges*	*(22093)*
	frequency	*%*	*frequency*	*%*
NL	688	3.11	11	0.05
IT	3236	14.65	66	0.30
∩ (NL,IT)	280	1.27	0	0.00

Table 27 Coverage of complete verb chains projected over Spanish wordnet structure

	nodes	*(2948)*	*edges*	*(2948)*
	frequency	*%*	*frequency*	*%*
NL	825	27.99	193	6.55
IT	968	32.84	221	7.50
∩ (NL,IT)	356	12.08	59	2.00

Table 28: Coverage of complete noun chains projected over Dutch wordnet structure

	nodes	*(50042)*	*edges*	*(50042)*
	frequency	*%*	*frequency*	*%*
ES	15062	30.10	3	0.01
IT	6882	13.75	1	0.00
∩ (ES,IT)	5188	10.37	1	0.00

Table 29: Coverage of complete verb chains projected over Dutch wordnet structure

	nodes	*(9965)*	*edges*	*(9965)*
	frequency	*%*	*frequency*	*%*
ES	2295	23.03	0	0.00
IT	1769	17.75	0	0.00
∩ (ES,IT)	882	8.85	0	0.00

Table 30: Coverage of complete noun chains projected over Italian wordnet structure

	nodes	*(173637)*	*edges*	*(173637)*
	frequency	*%*	*frequency*	*%*
ES	54063	31.14	60	0.03
NL	12784	7.36	44	0.03
∩ (ES,NL)	9740	5.61	2	0.00

Table 31: Coverage of complete verb chains projected over Italian wordnet structure

	nodes	*(63631)*	*edges*	*(63631)*
	frequency	*%*	*frequency*	*%*
ES	5976	9.39	8	0.01
NL	550	0.86	4	0.01
∩ (ES,NL)	276	0.43	0	0.00

Appendix II Projection of partial chains on the Dutch, Italian and Spanish wordnets

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

LENGTH	NL	IT	∩ (NL, IT)	ES
1	21980	22091	21868	22093
2	19852	21738	18999	22093
3	11746	20332	8405	22046
4	6101	12435	3566	21422
5	2572	9466	1361	19731
6	1041	6556	517	16113
7	313	3495	125	11139
8	57	1574	10	6178
9	4	543		3042
10		221		1408
11		39		519
12		11		198

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

LENGTH	NL	IT	∩ (NL, IT)	ES
1	9926	19231	5339	22093
2	2049	10200	427	22046
3	276	3579	38	21422
4	46	499	3	19731
5	1	36		16113
6		8		11139

Table 34: Coverage of partial VERB chains of NODES projected over Spanish wordnet structure

LENGTH	NL	IT	∩ (NL, IT)	ES
1	2771	2780	2608	2948
2	1810	1982	1343	2777
3	808	1072	489	1979
4	265	516	141	1096
5	57	196	23	503
6	10	58	6	205
7	2	12	2	80
8		2		30
9		1		3

Table 35: Coverage of partial VERB chains of EDGES projected over Spanish wordnet structure

LENGTH	NL	IT	∩ (NL, IT)	ES
1	801	1001	248	2777
2	116	139	8	1979
3	3	12		1096
4		1		503

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

LENGTH	ES	IT	∩ (ES, IT)	NL
1	50040	49977	49976	50042
2	49483	49625	49211	50041
3	45801	42492	41118	49960
4	38119	32858	30635	48960
5	26388	16664	14622	43696
6	15376	7029	6121	31231
7	7727	2416	2067	18574
8	3275	656	528	9095
9	1152	139	110	3581

10	348	11	6	1278
11	88	1	1	362
12	31			111
13	4			25

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

LENGTH	ES	IT	∩ (ES, IT)	NL
1	21343	21343	21343	50041
2	4350	4350	4350	49960
3	947	947	947	48960
4	156	156	156	43696

Table 38: Coverage of partial verb chains of NODES projected over Dutch wordnet structure

LENGTH	ES	IT	∩ (ES, IT)	NL
1	9965	9965	9965	9965
2	9901	9937	9844	9965
3	8934	9356	8656	9949
4	7821	7510	6611	9731
5	6174	4393	3696	8942
6	4375	2704	2233	7047
7	2833	1493	1178	4866
8	1479	774	572	3057
9	627	389	235	1741
10	231	156	58	814
11	67	41	11	306
12	8	4	1	100

Table 39: Coverage of partial verb chains of EDGES projected over Dutch wordnet structure

LENGTH	ES	IT	∩ (ES, IT)	NL
1	2768	1977	861	9965
2	167	96	8	9949
3	1	2		9731

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

LENGTH	ES	NL	∩ (ES, NL)	IT
1	173381	172497	172205	63631
2	167626	142224	136578	63631
3	141145	78287	69980	63468
4	108018	37831	31251	62431
5	77006	14303	11301	60171
6	46145	5238	4364	51478
7	26124	2137	1834	36669
8	16045	696	578	21698
9	10574	190	130	10948
10	3895			4688
11	261			960

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

LENGTH	ES	NL	∩ (ES, NL)	IT
1	74562	31932	18371	63631
2	23848	5281	1202	63468
3	7615	697	212	62431
4	916	35	2	60171
5	24	3		51478
6	1			36669

Table 42: Coverage of partial VERB chains of NODES projected over Italian wordnet structure

LENGTH	ES	NL	∩ (ES, NL)	IT
1	63631	63631	63631	63631
2	61231	62997	48194	63631
3	49800	21104	14156	63468
4	38621	7514	3483	62431
5	22158	1911	1097	60171
6	12004	281	214	51478
7	4570	10		36669
8	1644	2		21698
9	588			10948
10	28			4688

Table 43: Coverage of partial VERB chains of EDGES projected over Italian wordnet structure

LENGTH	ES	NL	∩ (ES, NL)	IT
1	12547	8113	4103	63631
2	2197	621	34	63468
	51	5		62431
3	4			60171

Table 44: Coverage of partial noun chains of NODES with 1 gap projected over Spanish wordnet structure

LENGTH	NL	IT	∩ (NL, IT)	ES
3	13752	3496	9331	22046
4	12449	3330	7738	21422
5	8876	2934	4924	19731
6	5495	2018	2638	16113
7	2502	1372	1018	11139
8	823	863	346	6178
9	217	471	81	3042
10	33	208	7	1408
11	1	83		519
12		43		198
13		7		87

Table 45: Coverage of partial noun chains of EDGES with 1 gap projected over Spanish wordnet structure

LENGTH	NL	IT	∩ (NL, IT)	ES
3	553	3330	205	21422
4	219	1830	78	19731
5	44	1326	13	16113
6	6	592	3	11139
7		51		6178

Table 46: Coverage of partial VERB chains of NODES with 1 gap projected over Spanish wordnet structure

LENGTH	NL	IT	∩ (NL, IT)	ES
3	410	277	352	1979
4	249	152	174	1096
5	90	80	43	503
6	22	39	11	205
7	10	20	4	80
8		4		30
9		1		3

Table 47: Coverage of partial VERB chains of EDGES with 1 gap projected over Spanish wordnet structure

LENGTH	NL	IT	∩ (NL, IT)	ES
3	30	10	0	1096
4	2	3		503

Table 48: Coverage of partial noun chains of NODES with 1 gap projected over Dutch wordnet structure

LENGTH	ES	IT	∩ (ES, IT)	NL
3	14810	15138	15498	49960
4	14345	14813	14799	48960
5	12395	9620	9103	43696
6	9245	6388	5880	31231
7	5677	3166	2869	18574
8	2638	1124	1010	9095
9	1029	326	273	3581
10	324	70	52	1278
11	67	8	5	362
12	22	1	1	111
13	5			25

Table 49: Coverage of partial noun chains of EDGES with 1 gap projected over Dutch wordnet structure

LENGTH	ES	IT	∩ (ES, IT)	NL
3	849	849	849	48960
4	164	164	164	43696
5	27	27	27	31231

Table 50: Coverage of partial verb chains of NODES with 1 gap projected over Dutch wordnet structure

LENGTH	ES	IT	∩ (ES, IT)	NL
3	1758	3113	3261	9949
4	1705	3104	3172	9731
5	1349	2916	2762	8942
6	947	2356	2089	7047
7	619	1686	1469	4866
8	379	955	738	3057
9	239	396	275	1741
10	101	159	98	814
11	32	67	35	306
12	14	22	1	100
13	7	4		29

Table 51: Coverage of partial verb chains of EDGES with 1 gap projected over Dutch wordnet structure

LENGTH	ES	IT	∩ (ES, IT)	NL
3	315	30	1	9731
4	59	2		8942
5	1			7047

Table 52: Coverage of partial noun chains of NODES with 1 gap projected over Italian wordnet structure

LENGTH	ES	NL	∩ (ES, NL)	IT
3	19249	83009	84453	169975
4	17977	59242	58122	148631
5	11578	33437	31340	111656
6	6804	17474	14799	72764
7	4635	7776	6535	47142
8	1891	2541	2224	23272
9	1279	689	570	16427
10	702	114	78	8554
11	119			1711

Table 53: Coverage of partial noun chains of EDGES with 1 gap projected over Italian wordnet structure

LENGTH	ES	NL	∩ (ES, NL)	IT
3	10293	383	55	148631
4	8077	29	1	111656
5	2537	1		72764
6	148			47142
7	1			23272

Table 54: Coverage of partial VERB chains of NODES with 1 gap projected over Italian wordnet structure

LENGTH	ES	NL	∩ (ES, NL)	IT
3	38301	51180	31266	63468
4	36005	50535	29119	62431
5	26424	37643	15773	60171
6	17748	19003	6953	51478
7	8759	6728	1583	36669
8	4038	1620	472	21698
9	1560	240	106	10948
10	484			4688

Table 55: Coverage of partial VERB chains of EDGES with 1 gap projected over Italian wordnet structure

LENGTH	ES	NL	∩ (ES, NL)	IT
3	796	93	0	62431
4	112			60171