

Compared and restructured wordnets for BCs in French, German, Czeck & Estonian

Version 1, Final

March 1st, 1999

Contributors:

Piek Vossen, University of Amsterdam

Claudia Kunze, Andrea Wagner, University of Tübingen

Karel Pala, University of Masarky

Kadri Vider, University of Tartu

LaurentCatherin, University of Avignon



**Deliverable 2D008
EuroWordNet, LE2-4003**

Identification number	LE-4003-D-017, D-034, D-036
Type	Document and Lingware
Title	Compared and restructured wordnets for BCs in French, German, Czeck & Estonian
Status	Final
Deliverable	2D008
Work Package	WP5
Task	n.a.
Period covered	December 1998 – March 1999
Date	March 1, 1999
Version	1
Number of pages	35
Authors	<ul style="list-style-type: none"> • Piek Vossen, Laura Bloksma, University of Amsterdam • Claudia Kunze, Andrea Wagner, University of Tübingen • Karel Pala, University of Masarky • Kadri Vider, University of Tartu • LaurentCatherin, University of Avignon
WP/Task responsible	AVI/MEM
Project contact point	<p>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@let.uva.nl</p>
EC project officer	Ray Hudson
Status	Public
Actual distribution	Project Consortium, the EuroWordNet User Group, the Ad Hoc Ansii Committee for Ontology standards, the world via http://www.let.uva.nl/~ewn .
Supplementary notes	n.a.

Key words	Linguistic Resources, Ontologies, Multilingual Wordnets, Language Engineering
Abstract	this deliverable we will describe the comparison of the core wordnets in EuroWordNet-2. The wordnets as such are described in 2D007, as well as the methodology that has been applied to build them. The result of the comparison is used to direct the finalization of the wordnets.
Status of the abstract	Complete
Received on	
Recipient's catalogue number	

Introduction

In this deliverable we will describe the comparison of the core wordnets in EuroWordNet-2. The wordnets as such are described in 2D007, as well as the methodology that has been applied to build them. The result of the comparison is used to direct the finalization of the wordnets. Another, similar, comparison will be done at the end of the project for the final wordnets. The comparison involves:

- an in-depth comparison of specific semantic fields using the Polaris database (section 2).
- a comparison of the Inter-Lingual-Index records that occur as translations of the synsets and the distribution of these ILI-records over the EuroWordNet top-ontology, as it has been added to WordNet1.5 (section 3).
- the compatibility of the hyponymy relations across the wordnets, represented by the ILI-record equivalents (section 4).

The comparisons are explained in detail in D014D015 and D027D028.

Table of Contents

I.	COMPARISON OF EWN2 IN POLARIS	6
1.	POLARIS COMPARISON FOR THE FRENCH WORDNET	7
1.1.	<i>Tables for summarising the results</i>	7
1.1.1.	Container.....	7
1.1.2.	Covering.....	8
1.1.3.	Cooking.....	9
2.	POLARIS COMPARISON FOR THE GERMAN WORDNET	FOUT! BLADWIJZER NIET GEDEFINIEERD.
2.1.	<i>Tables for summarising the results</i>	<i>Fout! Bladwijzer niet gedefinieerd.</i>
3.	POLARIS COMPARISON FOR THE ESTONIAN WORDNET	13
3.1.	<i>Tables for summarizing the results</i>	13
3.1.1.	Feelings.....	13
3.1.2.	Comestible	15
3.1.3.	Construction	16
4.	POLARIS COMPARISON FOR THE CZECH WORDNET	17
4.1.	<i>Tables for summarizing the results</i>	17
II.	CONCLUSION ON THE POLARIS COMPARISON	19
III.	INTERSECTION OF ILI RECORDS AND TOP-ONTOLOGY CLUSTERING	20
1.	CONCLUSIONS:	21
IV.	CONCLUSIONS:	30
V.	COMPARISON OF ILI-CHAINS IN EWN2	31
1.	EVALUATION OF INDIVIDUAL WORDNETS	32
2.	CONCLUSIONS:	35

I. Comparison of EWN2 in Polaris

Each EWN2 partner has compared 3 semantic clusters using Polaris:

TAR (Estonian):	Building; Comestible; Feelings
AVI (French):	Container; Covering; Cooking
TUE (German):	Garment; Covering; Movement
BRN (Czech):	Instrument; Vehicle; Sound

All groups have circulated the most representative hyperonyms or tops for these 12 fields. Comparison in Polaris is roughly carried out by:

1. expanding the hyponym trees for the cluster representatives in the other languages and your mother language. The list of hyponyms can be saved in Polaris as Favorites (either in the database or as external files).
2. project the clusters in the other languages to your own language.

Projecting the foreign-language clusters to your mother language gives information on the hyponyms which are shared, what hyponyms are unique in your mother language wordnet or are shared by all the others but not by your own wordnet. The other wordnets are then used as a so-called Reference wordnet to evaluate and inspect your own wordnet. Inspection of the differences are summarized in a report by each site explaining the main problems/errors. Such a comparison will give information on:

1. *the quality and quantity of equivalence relations*
2. *the overlap across wordnets*
3. *the coherence of classification*

1. Polaris comparison for the French wordnet

Areas compared are:

- ◆ Container, covering (FirstOrder)
- ◆ Cooking (SecondOrder)

Major nodes, hyponyms and equivalence relations

- ◆ French wordnet does not have equivalence relations other than eq_synonym
- ◆ Due to the method the French wordnet correspondance to ILIs is one to one, so number of french synset equals number of ILIs

Projections of reference WNs to French

- ◆ French Wordnet covers around the half of the synset projected from other Wordnets.

Comparing Projections

- ◆ Due to the method, translating Wordnet 1.5 to French, we couldn't have a lot of

The next table give an overview over the results of comparing semantic clusters for container, covering and cooking. There is no result for the Czeck wordnet because they don't have actualy produced internal relations.

1.1. Tables for summarising the results

1.1.1. Container

Table 1 : Projection of Clusters for the field Container to the French wordnet

<i>Wordnet</i>	<i>No. Synsets in cluster</i>	<i>No. Of Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Matched Target ILIs</i>
GE	54	40	18	22
EE	28	27	13	14
CZ	X	X	X	X
FR	52	52	X	X
Union	X	54	27	27

Table 2 : Comparison of Projected Container Clusters with French Cluster (52 WMs)

<i>Wordnet</i>	<i>Matched Target ILIs</i>	<i>Projected Target WMs</i>	<i>Union</i>	<i>Intersection</i>	<i>Unique WMs in Projected Cluster</i>	<i>Unique WMs in Target Cluster</i>
GE	22	22	54	20	2	32
EE	14	14	52	14	0	38
CZ	X	X	X	X	X	X
Union	54	27	54	25	2	27

Table 3 : Types of mismatches

<i>Type of Mismatch</i>	<i>Number of WMs</i>
Different classification in the French wordnet	2
Wrong classification in the French wordnet	0
wrong translation in the French wordnet	0

1.1.2. Covering**Table 4 : Projection of Clusters for the field Covering to the French wordnet**

<i>Wordnet</i>	<i>No. synsets in cluster</i>	<i>No. Of Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Matched Target ILIs</i>
GE	68	41	15	26
EE	33	30	9	21
CZ	X	X	X	X
FR	70	70	X	X
Union	X	55	23	32

Table 5 : Comparison of Projected Covering Clusters with French Cluster (70 WMs)

<i>Wordnet</i>	<i>Matched Target ILIs</i>	<i>Projected Target WMs</i>	<i>Union</i>	<i>Intersection</i>	<i>Unique WMs in Projected Cluster</i>	<i>Unique WMs in Target Cluster</i>
GE	26	26	73	23	3	47
EE	21	21	70	21	0	49
CZ	X	X	X	X	X	X
Union	32	32	73	29	3	41

Table 6 : Types of mismatches

<i>Type of Mismatch</i>	<i>Number of WMs</i>
Different classification in the French wordnet	3
Wrong classification in the French wordnet	0
Wrong translation in the French wordnet	0

1.1.3. Cooking

Table 7 : Projection of Clusters for the field Container to the French wordnet

<i>Wordnet</i>	<i>No. synsets in cluster</i>	<i>No. of Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Matched Target ILIs</i>
GE	4	4	2	2
EE	50	32	21	11
CZ	X	X	X	X
FR	7	7	X	X
Union	50	35	23	12

Table 8 : Comparison of Projected Container Clusters with French Cluster (365 WMs)

<i>Wordnet</i>	<i>Matched Target ILIs</i>	<i>Projected Target WMs</i>	<i>Union</i>	<i>Intersection</i>	<i>Unique WMs in Projected Cluster</i>	<i>Unique WMs in Target Cluster</i>
GE	2	2	10	2	0	8
EE	11	11	13	8	3	2
CZ	X	X	X	X	X	X
Union	12	12	13	9	3	1

Table 9 : Types of mismatches

<i>Type of Mismatch</i>	<i>Number of WMs</i>
Different classification in the French wordnet	3
Wrong classification in the French wordnet	0
wrong translation in the French wordnet	0

2. Comparison of Semantic Clusters for Garment, Movement and Place.

This report gives an overview over the results of comparing semantic clusters for Garment, Movement and Place. We employed the main steps for comparison distributed by Amsterdam.

As starting point, we took the following hyperonyms into consideration:

Czech: garment: oblecení 1 (could not be explored in the database)
 movement: pohyb 1, pohyb 5 (the latter reading was not available)
 place: místo 4, místo 1

Since the Czech wordnet so far has not provided for language internal relations, we explored the Czech hyponyms via projection to WordNet 1.5 concepts.

German: garment: Bekleidung 1
 movement: Bewegung 2, Bewegung 3, bewegen 1
 place: Ort 1, befinden 2, wohnen 1

Estonian: garment: riie 1
 movement: liikumine 1
 place: koht 4

French: garment: vêtement 1
 movement: mouvement 2, mouvement 4, bouger 3
 place: emplacement 1

2.1. Tables for summarising the results

2.1.1. Garment

Table 10 : Projections of Clusters for the field Garment to the German wordnet

<i>Wordnet</i>	<i>No. of synsets in cluster</i>	<i>No. of Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Matched Target ILIs</i>
CZ	0	0	0	0
DE	52	35	x	x
EE	12	10	2	8
FR	40	40	21	19
Union	x	42	23	19

Table 11 : Comparison of Projected Garment Clusters with German Cluster (52 WMs)

<i>Word net</i>	<i>Matched Target ILIs</i>	<i>Projected Target ILIS</i>	<i>Union</i>	<i>Intersection</i>	<i>Unique WMs in Projected Cluster</i>	<i>Unique WMs in Target Cluster</i>
CZ	0	0	52	0	0	52
EE	8	11	52	11	0	41
FR	19	33	52	33	0	19
Union	27	33	52	33	0	19

Table 12 : Mismatches in Semantic Cluster Garment

<i>Type of Mismatch</i>	<i>Number of WMs</i>
Different classification in the German wordnet	0
Wrong classification in the German wordnet	0
Wrong translation in the German wordnet	0

Alternative classification/Cross Classification: Covering

2.1.2. Movement

Table 13 : Projections of Clusters for the field Movement to the German wordnet

<i>Wordnet</i>	<i>No. of synsets in cluster</i>	<i>No. of Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Matched Target ILIs</i>
CZ	26	26	22	4
DE	368	237	x	x
EE	7	7	2	5
FR	217	217	158	59
Union	X	247	181	66

Table 14 : Comparison of Projected Movement Clusters with German Cluster (367 WMs)

<i>Word net</i>	<i>Matched Target ILIs</i>	<i>Projected Target ILIS</i>	<i>Union</i>	<i>Intersection</i>	<i>Unique WMs in Projected Cluster</i>	<i>Unique WMs in Target Cluster</i>
CZ	4	4	325	1	3	321
EE	5	8	327	3	5	319
FR	59	114	392	44	70	278
Union	69	124	399	47	77	275

Table 15 : Mismatches in Semantic Cluster Movement

<i>Type of Mismatch</i>	<i>Number of WMs</i>
Different classification in the German wordnet	0
Wrong classification in the German wordnet	8
Wrong translation in the German wordnet	2

Alternative classification/Cross Classification: Event, Happening

2.1.3. Place

Table 16 : Projections of Clusters for the field Place to the German wordnet

<i>Wordnet</i>	<i>No. Of synsets in cluster</i>	<i>No. of Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Matched Target ILIs</i>
CZ	4	4	2	2
DE	367	356	x	x
EE	117	68	34	34
FR	163	158	101	57
Union	X	190	126	64

Table 17 : Comparison of Projected Place Clusters with German Cluster (368 WMs)

<i>Word net</i>	<i>Matched Target ILIs</i>	<i>Projected Target ILIS</i>	<i>Union</i>	<i>Intersection</i>	<i>Unique WMs in Projected Cluster</i>	<i>Unique WMs in Target Cluster</i>
CZ	2	2	368	1	1	366
EE	34	35	371	31	4	336
FR	57	58	376	49	9	318
Union	93	62	377	52	10	315

Table 18 : Mismatches in Semantic Cluster Place

<i>Type of Mismatch</i>	<i>Number of WMs</i>
Different classification in the German wordnet	2
Wrong classification in the German wordnet	29
Wrong translation in the German wordnet	0

Alternative classification/Cross Classification: Abstract Location, Possession

Main reason for high number of wrong classifications within German cluster Place is the local interpretation of (thematical) *domain* (*Bereich* 2 which means *field* 15). That's why a lot of hyponyms are erroneously classified.

2.1.4. conclusion

Table 19 : Types of mismatches (Summary)

<i>Type of Mismatch</i>	<i>Number of WMs</i>
Different classification in the German wordnet	8
Wrong classification in the German wordnet	37
Wrong translation in the German wordnet	2

Example for different classification: Several «movement» concepts have been classified as events within the German wordnet, for instance *gliding 1* (*Segelfliegen 1*), *journey 1* (*Reise 1*) and similar concepts.

Example for wrong translation: *arrange 4* has been wrongly translated to *organisieren 5* which means something like *arrange 6*.

Example for wrong classification: The German concept *anziehen 1* (*get_dressed 1*) is classified as a movement verb.

3. Polaris comparison for the Estonian wordnet

3.1. Tables for summarizing the results

3.1.1. Feelings

Table 20 : Projection of Clusters for the field Feeling to the Estonian wordnet (Noun)

<i>Wordnet</i>	<i>No. synsets in cluster</i>	<i>No. of Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Matched Target ILIs</i>
EE	13	12		
DE	139	126	102	24
FR	77	77	58	19
CZ				
Union	x	165	135	30

Table 21 : Projection of Clusters for the field Feeling to the Estonian wordnet (Verb)

<i>Wordnet</i>	<i>No. synsets in cluster</i>	<i>No. of Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Matched Target ILIs</i>
EE	4	5		
DE	85	63	49	14
FR	19	19	14	5
CZ				
Union	x	75	59	16

Table 22 : Comparison of Projected Feeling Clusters with Estonian Cluster (13 WMs) (Noun)

<i>Wordnet</i>	<i>Matched Target ILIs</i>	<i>Projected Target WMs</i>	<i>Union</i>	<i>Intersection</i>	<i>Unique WMs in Projected Cluster</i>	<i>Unique WMs in Target Cluster</i>
CZ						
DE	24	25	30	8	17	5
FR	19	20	23	10	10	3
Union	30	31	34	10	21	3

Table 23 : Comparison of Projected Feeling Clusters with Estonian Cluster (4 WMs) (Verb)

<i>Wordnet</i>	<i>Matched Target ILIs</i>	<i>Projected Target WMs</i>	<i>Union</i>	<i>Intersection</i>	<i>Unique WMs in Projected Cluster</i>	<i>Unique WMs in Target Cluster</i>
CZ						
DE	14	16	17	3	13	1
FR	5	7	8	3	4	1
Union	16	18	18	4	14	0

Table 24 : Types of mismatches

<i>Type of Mismatch</i>	<i>Number of WMs</i>	<i>Noun</i>	<i>Verb</i>
Different classification in the Estonian wordnet		13	5
Wrong classification in the Estonian wordnet		8	5
wrong translation in the Estonian wordnet (a Estonian word which is not a feeling is translated to a feeling in WN1.5)			

Wrong classification in the Estonian among feeling-verbs is caused by fact, that we have not classified these verbs into hyperonym/hyponym trees yet.

3.1.2. Comestible

Table 25 : Projection of Clusters for the field Comestible to the Estonian wordnet (Noun)

<i>Wordnet</i>	<i>No. synsets in cluster</i>	<i>No. of Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Matched Target ILIs</i>
EE	235	141		
DE	357	257	187	70
FR	214	204	135	69
CZ				
Union	x	383	289	94

Table 26 : Projection of Clusters for the field Comestible to the Estonian wordnet (Verb)

<i>Wordnet</i>	<i>No. synsets in cluster</i>	<i>No. of Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Matched Target ILIs</i>
EE	10	7		
DE	51	31	23	8
FR				
CZ				
Union	x	31	23	8

Table 27 : Comparison of Projected Comestible Clusters with Estonian Cluster (235 WMs) (Noun)

<i>Wordnet</i>	<i>Matched Target ILIs</i>	<i>Projected Target WMs</i>	<i>Union</i>	<i>Intersection</i>	<i>Unique WMs in Projected Cluster</i>	<i>Unique WMs in Target Cluster</i>
CZ						
DE	70	80	86	32	48	6
FR	69	81	83	36	45	2
Union	94	106	107	37	69	1

Table 28 : Comparison of Projected Comestible Clusters with Estonian Cluster (235 WMs) (Verb)

<i>Wordnet</i>	<i>Matched Target ILIs</i>	<i>Projected Target WMs</i>	<i>Union</i>	<i>Intersection</i>	<i>Unique WMs in Projected Cluster</i>	<i>Unique WMs in Target Cluster</i>
CZ						
DE	8	11	11	11	0	0
FR						
Union	8	11	11	11	0	0

Table 29 : Types of mismatches (69)

<i>Type of Mismatch</i>	<i>Number of WMs</i>
Different classification in the Estonian wordnet	49
Wrong classification in the Estonian wordnet	5
wrong translation in the Estonian wordnet (a Estonian word which is not a comestible is translated to a comestible in WN1.5)	

The rest of unmatches are hyponyms for:

- meal – the flesh of animals (including fishes and birds and snails) used as food
- soup - liquid food esp. of meat or fish or vegetable stock often containing spieces of solid food
- dairy product, hyponyms like curd, cream and yoghurt

and some concepts, lexicalized in Estonian, but have only eq_has_hyperonym relation to ILI – as Rhine wine via *wine*, *vino* and mixed spices via *spice*.

3.1.3. Construction

Table 30 : Projection of Clusters for the field Construction to the Estonian wordnet (Noun)

<i>Wordnet</i>	<i>No. synsets in cluster</i>	<i>No. of Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Matched Target ILIs</i>
EE	97	75	x	x
DE	60	49	25	24
FR	152	149	107	42
CZ				
Union		176	126	50

Table 31 : Comparison of Projected Construction Clusters with Estonian Cluster (97 WMs) (Noun)

<i>Wordnet</i>	<i>Matched Target ILIs</i>	<i>Projected Target WMs</i>	<i>Union</i>	<i>Intersection</i>	<i>Unique WMs in Projected Cluster</i>	<i>Unique WMs in Target Cluster</i>
CZ						
DE	24	28	82	26	2	54
FR	42	48	85	43	5	37
Union	50	56	86	50	6	30

Table 32 : Types of mismatches

<i>Type of Mismatch</i>	<i>Number of WMs 3</i>
Different classification in the Estonian wordnet	1
Wrong classification in the Estonian wordnet	4
wrong translation in the Estonian wordnet (a Estonian word which is not a building is translated to a building in WN1.5)	0

For ILI Base Concepts that have no lexicalization in Estonian we have created dummy synsets *SEMAUK*. Among the topic *Construction* there was one such: *place of business 1, business establishment 1*. That accounts for the one non-classified mismatch in Table 3.

4. **Polaris comparison for the Czech wordnet**

4.1. Tables for summarizing the results

Table 33 : Projection of the Clusters from Czech to other WNs: Measuring Instruments

<i>Wordnet</i>	<i>Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Target ILIs</i>	
CZ-ESTO	62	17	45	
CZ-DE	62	58	4	
CZ-FR	62	52	12	
CZ-NED	62	29	33	
CZ-ITA	62	38	24	
CZ-ES	62	35	27	
CZ-WN 1.5	62	0	62	

Table 34 : Projection of the Clusters: Musical Instruments

<i>Wordnet</i>	<i>Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Target ILIs</i>	
CZ-ESTO	39	1	38	
CZ-DE	39	14	25	
CZ-FR	39	32	7	
CZ-NED	39	3	36	
CZ-ITA	39	13	26	
CZ-ES	39	11	28	
CZ-WN 1.5	39	0	39	

Table 35 : Projection of the Clusters: Sounds

<i>Wordnet</i>	<i>Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Target ILIs</i>	
CZ-ESTO	41	1	38	
CZ-DE	41	14	25	
CZ-FR	41	37	4	
CZ-NED	41	3	36	
CZ-ITA	41	13	26	
CZ-ES	41	11	28	
CZ-WN 1.5	41	0	41	

Table 36 : Projection of the Clusters: Vehicle

<i>Wordnet</i>	<i>Source ILIs</i>	<i>Unmatched ILIs</i>	<i>Target ILIs</i>	
CZ-ESTO	187	174	13	
CZ-DE	187	166	21	
CZ-FR	187	152	35	
CZ-NED	187	148	39	
CZ-ITA	187	113	74	
CZ-ES	187	71	116	
CZ-WN 1.5	187	0	187	

Remark:

The French data have been downloaded again from ftp site, this time successfully. The complete match with WN 1.5 is due to the fact that it was used as a starting point. If we had more time the results would have been different since there are differences that had not been reflected in the present clusters. For example, in WN 1.5 we have not found strum (drnkací) musical instruments. Or in vehicles WN 1.5 does not differentiate tram and trolley-bus, in Czech trams are understood as running on rails and trolleybuses on roads.

Table 37 : Types of mismatches

Type of Mismatch	Number of WMs
Different classification in the Czech wordnet	9
Wrong classification the Czech wordnet	9
Wrong translation in in the Czech wordnet	14
Total	32

After the quick scanning the list it appears that most of the errors are wrong translations, eg. trafikant is a person selling tobacco, cigarettes and newspapers and not a shop selling these articles: for this a proper expression is trafika. Six cases of the wrong translations have its origin in wrong Czech synsets, where non synonymical expressions came together in one synset, eg. krokev and vora• have been put in one synset and the projection is right for the first expression (a sort of beam in the roof) while it does not fit for the second one -- vora• (rafter). There is an interesting example of the different classification, the concept pracovní tábor (labour camp) has obtained the following gloss: for trustworthy prisoners employed in government projects. That certainly cannot be said about the labour camps in the previous communist regime where they were rather gulags though the expression gulag is not now a common expression in the standard Czech.

Our data have been obtained via projection process only. Our strategy is that we would like to finish the preparation of the main body of the data for Czech Wordnet first and only then concentrate fully on hypero/hyponymical relations and ILR.

II. Conclusion on the Polaris Comparison

In this document we described the comparison between German, Czech, Estonian and French wordnets on specific clusters. This test is a bit distorted concerning the Czech Wordnet because of the lack of internal relations. There is no comparison between Estonian and Czech and between French and Czech. The comparison between German and Czech was done by projecting the Czech synsets over the Wordnet 1.5 hierarchy.

This comparison allows us to find the mistakes of classification which has been done. For each Wordnet the most part of the mistakes came from differences of classification, especially for the Estonian Wordnet. The German Wordnet has a higher number of errors of classification which are due to a problem of interpretation of a domain.

Table 38 : type of mismatches

<i>Language</i>	<i>Fields</i>	<i>Classification Difference</i>	<i>Classification Error</i>	<i>Translation Error</i>
FR	Container; Covering; Cooking	8		
DE	Garment; Covering; Movement	8	37	2
EE	Building; Comestible; Feelings	68	22	
CZ	Instrument; Vehicle; Sound	9	9	14

III. Intersection of ILI records and top-ontology clustering

The next 3 tables show the intersection of the ILI records that occur as translations of the first subset wordnets.

<i>Nouns</i>	<i>Total</i>	<i>62780</i>	<i>12995</i>
	Frequency	\cup (WN,DE,FR, EE, CZ)	\cup (DE, FR, EE, CZ)
DE	4480	7,1%	34,5%
FR	5523	8,8%	42,5%
EE	2594	4,1%	20,0%
CZ	6754	10,8%	52,0%
\cap (DE, FR)	1899	3,0%	14,6%
\cap (DE, EE)	1192	1,9%	9,2%
\cap (DE, CZ)	1556	2,5%	12,0%
\cap (FR, EE)	1399	2,2%	10,8%
\cap (FR, CZ)	1863	3,0%	14,3%
\cap (EE, CZ)	1097	1,7%	8,4%
\cap (DE, FR, EE)	955	1,5%	7,3%
\cap (DE, FR, CZ)	875	1,4%	6,7%
\cap (DE, EE, CZ)	656	1,0%	5,0%
\cap (FR, EE, CZ)	716	1,1%	5,5%
\cap (DE, FR, EE, CZ)	552	0,9%	4,2%

<i>Verbs</i>	<i>Total</i>	<i>12215</i>	<i>4135</i>
	Frequency	\cup (WN,DE, FR, EE, CZ)	\cup (DE, FR, EE, CZ)
DE	1959	16,0%	47,4%
FR	2534	20,7%	61,3%
EE	489	4,0%	11,8%
CZ	1306	10,7%	31,6%
\cap (DE, FR)	905	7,4%	21,9%
\cap (DE, EE)	328	2,7%	7,9%
\cap (DE, CZ)	584	4,8%	14,1%
\cap (FR, EE)	355	2,9%	8,6%
\cap (FR, CZ)	621	5,1%	15,0%
\cap (EE, CZ)	327	2,7%	7,9%
\cap (DE, FR, EE)	283	2,3%	6,8%
\cap (DE, FR, CZ)	372	3,0%	9,0%
\cap (DE, EE, CZ)	203	1,7%	4,9%
\cap (FR, EE, CZ)	214	1,8%	5,2%
\cap (DE, FR, EE, CZ)	194	1,6%	4,7%

<i>Total</i>	<i>Total</i>	<i>74995</i>	<i>26392</i>
	Frequency	$\cup(\text{WN,DE, FR, EE, CZ})$	$\cup (\text{DE, FR, EE, CZ})$
DE	6439	8,6%	24,4%
FR	8057	10,7%	30,5%
EE	3083	4,1%	11,7%
CZ	8060	10,7%	30,5%
$\cap (\text{DE, FR})$	2804	3,7%	10,6%
$\cap (\text{DE, EE})$	1520	2,0%	5,8%
$\cap (\text{DE, CZ})$	2140	2,9%	8,1%
$\cap (\text{FR, EE})$	1754	2,3%	6,6%
$\cap (\text{FR, CZ})$	2484	3,3%	9,4%
$\cap (\text{EE, CZ})$	1424	1,9%	5,4%
$\cap (\text{DE, FR, EE})$	1238	1,7%	4,7%
$\cap (\text{DE, FR, CZ})$	1247	1,7%	4,7%
$\cap (\text{DE, EE, CZ})$	859	1,1%	3,3%
$\cap (\text{FR, EE, CZ})$	930	1,2%	3,5%
$\cap (\text{DE, FR, EE, CZ})$	746	1,0%	2,8%

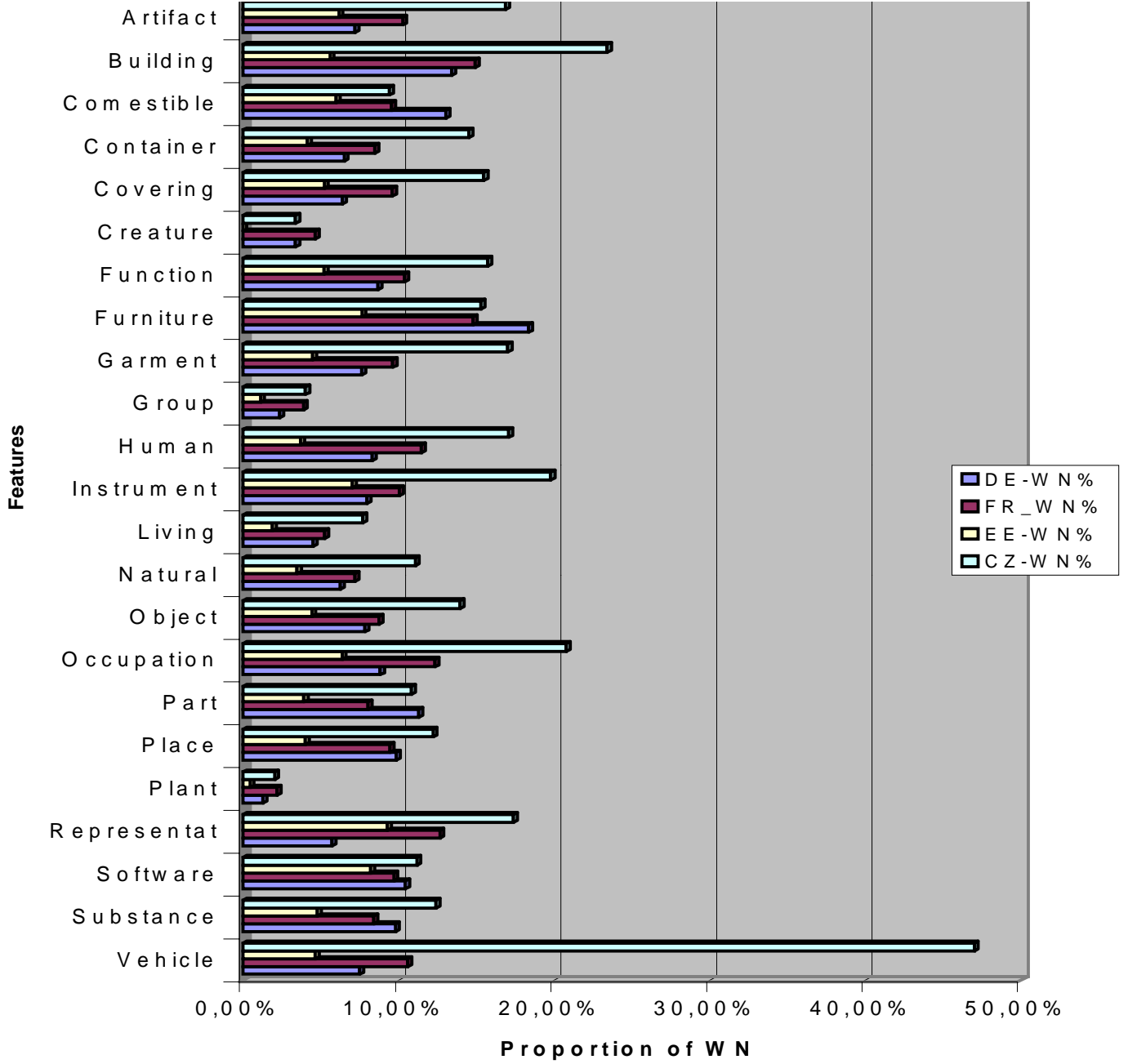
1. Conclusions:

The proportion of WordNet is between 4 and 10% for nouns and between 4 and 20% for verbs. This is good considering the fact that the total aim for French and German is about 15% of WordNet1.5 (7.5% for Czech and Estonian). The intersection of 3 languages ranges from 5 to 7,3% for nouns and from 4,9 to 9% for verbs, which is better than the first subset in EWN1. The intersection for 4 languages is obviously lower. These figures have to improve for the final subsets (compare 25% intersection for EWN1), although the limited size of the wordnets will make it more difficult to get higher percentages of intersection.

To evaluate the conceptual distribution of these index records we have imported the ILI-records into the ALS database and extracted the top-ontology classification for each record, via its hyperonym chain in WordNet1.5. The top-ontology has been added to by AMS to 793 Noun Synsets and 617 Verbs Synsets, including all Common Base Concepts and all remaining top-nodes (mainly verbs) that are not included in the set of Common Base Concepts. Furthermore, the WordNet1.5 Lexicographer's file codes have been included and can be used to compare the lists of ILI-records. The results are listed below in separate tables for concrete nouns, abstract nouns and for verbs.

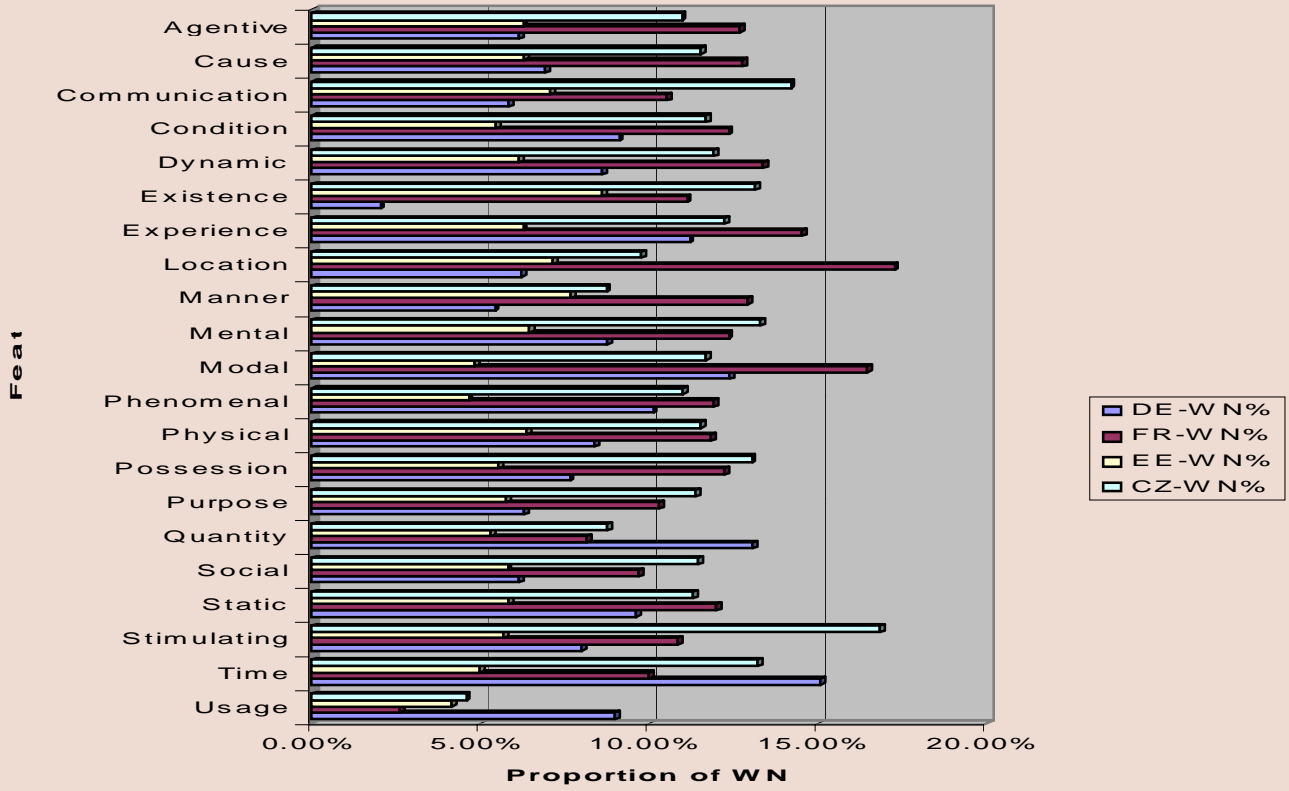
For each wordnet we give 3 columns. The first column gives the number TC synsets that inherited a certain Top Concept. The second column gives the percentage of the total set of inherited features, and the third column gives the percentage of the synsets in WordNet1.5 that inherited same feature. For Subset1 in EWN2, it is sufficient that 7.5% up to 10% of WordNet1.5 is covered.

	WN		DE		FR		EE		CZ		CZ-WN%	
	Syns	% of wn	DE	%DE	FR	%FR	FR_WN%	%EE	EE-WN%	CZ		%CZ
Animal	6958	3,67%	195	1,55%	142	0,99%	2,80%	70	1,00%	269	1,23%	3,87%
Artifact	12446	6,57%	899	7,13%	1279	8,88%	7,22%	769	11,04%	2101	9,59%	16,88%
Building	589	0,31%	79	0,63%	88	0,61%	13,41%	33	0,47%	138	0,63%	23,43%
Comestible	2304	1,22%	301	2,39%	220	1,53%	13,06%	138	1,98%	217	0,99%	9,42%
Container	1060	0,56%	69	0,55%	90	0,62%	6,51%	44	0,63%	154	0,70%	14,53%
Covering	1279	0,67%	82	0,65%	123	0,85%	6,41%	67	0,96%	198	0,90%	15,48%
Creature	473	0,25%	16	0,13%	22	0,15%	3,38%		0,00%	16	0,07%	3,38%
Function	21284	11,23%	1847	14,65%	2213	15,36%	8,68%	1110	15,93%	3349	15,28%	15,73%
Furniture	196	0,10%	36	0,29%	29	0,20%	18,37%	15	0,22%	30	0,14%	15,31%
Garment	446	0,24%	34	0,27%	43	0,30%	7,62%	20	0,29%	76	0,35%	17,04%
Group	13113	6,92%	309	2,45%	509	3,53%	2,36%	150	2,15%	529	2,41%	4,03%
Human	6862	3,62%	570	4,52%	787	5,46%	8,31%	255	3,66%	1172	5,35%	17,08%
Instrument	4557	2,40%	363	2,88%	459	3,19%	7,97%	320	4,59%	902	4,12%	19,79%
Living	23704	12,51%	1069	8,48%	1246	8,65%	4,51%	446	6,40%	1827	8,34%	7,71%
Natural	37427	19,75%	2343	18,58%	2702	18,76%	6,26%	1305	18,73%	4149	18,93%	11,09%
Object	27603	14,57%	2163	17,16%	2414	16,76%	7,84%	1223	17,55%	3854	17,59%	13,96%
Occupation	1222	0,64%	108	0,86%	151	1,05%	8,84%	78	1,12%	254	1,16%	20,79%
Part	7412	3,91%	838	6,65%	597	4,14%	11,31%	291	4,18%	802	3,66%	10,82%
Place	3253	1,72%	321	2,55%	307	2,13%	9,87%	131	1,88%	398	1,82%	12,23%
Plant	8221	4,34%	107	0,85%	181	1,26%	1,30%	38	0,55%	168	0,77%	2,04%
Representat	592	0,31%	34	0,27%	75	0,52%	5,74%	55	0,79%	103	0,47%	17,40%
Software	134	0,07%	14	0,11%	13	0,09%	10,45%	111	0,16%	15	0,07%	11,19%
Substance	7912	4,18%	776	6,16%	665	4,62%	9,81%	377	5,41%	981	4,48%	12,40%
Vehicle	453	0,24%	34	0,27%	48	0,33%	7,51%	21	0,30%	213	0,97%	47,02%
Total	189500		12607		14403		6,65%	6967		21915		11,56%

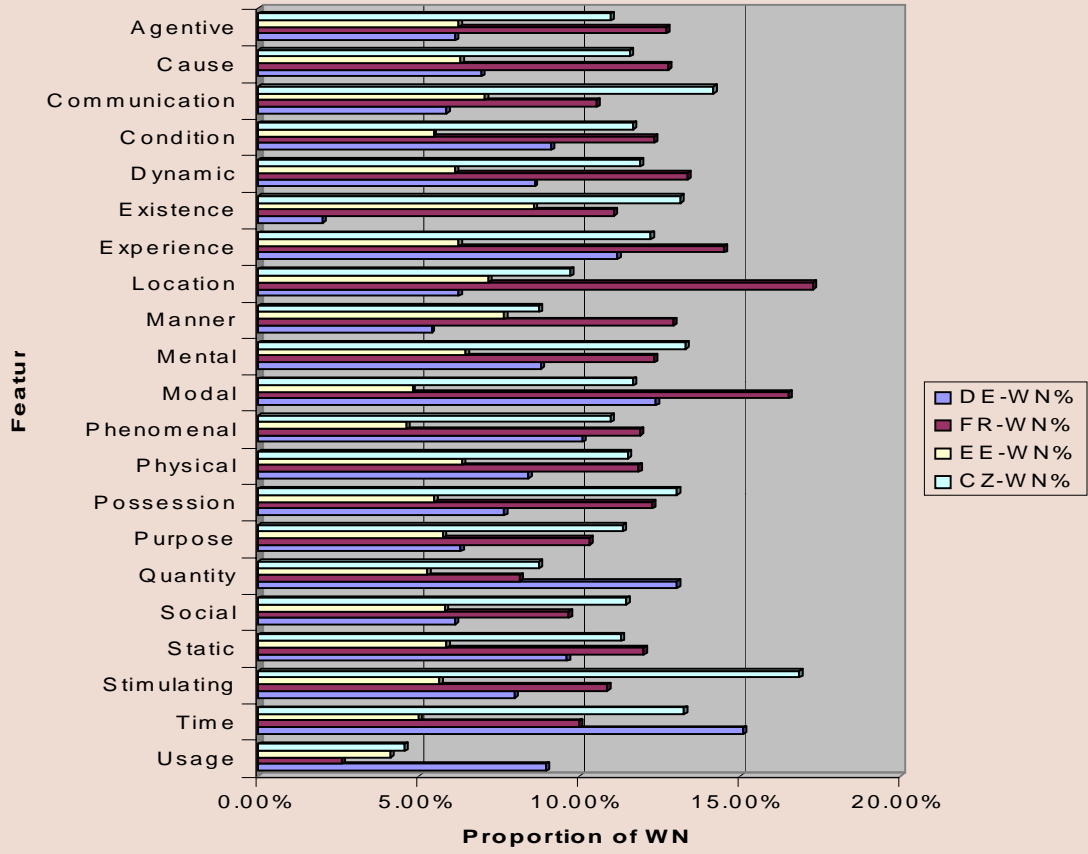


Top-Concept	WN		DE		FR		EE		CZ		CZ-WN%	
	Syns	% of wn	DE	%DE	FR	%FR	FR_WN %	EE	%EE	EE-WN%		CZ
Agentive	7214	6,80%	444	5,20%	917	7,22%	12,71%	452	7,00%	6,27%	791	6,39%
BoundedEvent	4753	4,48%	343	4,02%	611	4,81%	12,86%	322	4,99%	6,77%	555	4,48%
Cause	9071	8,56%	630	7,38%	1158	9,11%	12,77%	571	8,84%	6,29%	1049	8,47%
Communication	4325	4,08%	253	2,96%	457	3,60%	10,57%	305	4,72%	7,05%	614	4,96%
Condition	2325	2,19%	212	2,48%	287	2,26%	12,34%	127	1,97%	5,46%	272	2,20%
Dynamic	11760	11,09%	1013	11,86%	1574	12,39%	13,38%	721	11,17%	6,13%	1400	11,31%
Existence	198	0,19%	4	0,05%	22	0,17%	11,11%	17	0,26%	8,59%	26	0,21%
Experience	4012	3,78%	449	5,26%	583	4,59%	14,53%	251	3,89%	6,26%	491	3,97%
Location	851	0,80%	53	0,62%	147	1,16%	17,27%	61	0,94%	7,17%	83	0,67%
Manner	573	0,54%	31	0,36%	74	0,58%	12,91%	44	0,68%	7,68%	50	0,40%
Mental	6275	5,92%	551	6,45%	775	6,10%	12,35%	405	6,27%	6,45%	836	6,75%
Modal	291	0,27%	36	0,42%	48	0,38%	16,49%	14	0,22%	4,81%	34	0,27%
Phenomenal	1227	1,16%	124	1,45%	146	1,15%	11,90%	57	0,88%	4,65%	135	1,09%
Physical	4714	4,45%	396	4,64%	559	4,40%	11,86%	300	4,65%	6,36%	544	4,39%
Possession	889	0,84%	68	0,80%	109	0,86%	12,26%	49	0,76%	5,51%	116	0,94%
Property	6999	6,60%	761	8,91%	970	7,63%	13,86%	399	6,18%	5,70%	761	6,15%
Purpose	9250	8,73%	582	6,81%	953	7,50%	10,30%	534	8,27%	5,77%	1052	8,50%
Quantity	2228	2,10%	291	3,41%	182	1,43%	8,17%	118	1,83%	5,30%	195	1,58%
Relation	4154	3,92%	257	3,01%	426	3,35%	10,26%	265	4,10%	6,38%	566	4,57%
Social	7449	7,03%	459	5,37%	721	5,67%	9,68%	432	6,69%	5,80%	853	6,89%
Static	12522	11,81%	1203	14,09%	1505	11,84%	12,02%	733	11,35%	5,85%	1418	11,45%
Stimulating	599	0,57%	48	0,56%	65	0,51%	10,85%	34	0,53%	5,68%	101	0,82%
Time	822	0,78%	124	1,45%	82	0,65%	9,98%	41	0,64%	4,99%	109	0,88%
UnboundedEvent	2792	2,63%	143	1,67%	316	2,49%	11,32%	174	2,70%	6,23%	296	2,39%
Usage	723	0,68%	65	0,76%	19	0,15%	2,63%	30	0,46%	4,15%	33	0,27%
Total	106016		8540		12706		11,98%	6456		6,09%	12380	

EWN2 Abstract Nouns



EWN2 Verbs



WN Lexicographer's file code		WN		DE		FR		EE		CZ		
		Syns	% of wn	DE	%DE	FR	%FR	FR_WN	%EE	EE-WN%	CZ	%CZ
4	noun.act	4953	7.13%	281	6.19%	614	11.31%	290	11.29%	492	6.94%	9.9%
5	noun.animal	6742	9.70%	168	3.70%	120	2.21%	57	2.22%	239	3.37%	3.5%
6	noun.artifact	9420	13.56%	676	14.90%	983	18.10%	578	22.50%	1741	24.55%	18.5%
7	noun.attribute	2553	3.67%	231	5.09%	367	6.76%	158	6.15%	183	2.58%	7.2%
8	noun.body	1692	2.43%	195	4.30%	124	2.28%	67	2.61%	210	2.96%	12.4%
9	noun.cognition	2409	3.47%	258	5.69%	329	6.06%	177	6.89%	337	4.75%	14.0%
10	noun.communication	4144	5.96%	230	5.07%	425	7.83%	280	10.90%	588	8.29%	14.2%
11	noun.event	787	1.13%	66	1.45%	110	2.03%	24	0.93%	136	1.92%	17.3%
12	noun.feeling	361	0.52%	61	1.34%	54	0.99%	8	0.31%	31	0.44%	8.6%
13	noun.food	2293	3.30%	295	6.50%	212	3.90%	133	5.18%	212	2.99%	9.2%
14	noun.group	7273	10.47%	151	3.33%	257	4.73%	75	2.92%	261	3.68%	3.6%
15	noun.location	1889	2.72%	218	4.80%	153	2.82%	56	2.18%	177	2.50%	9.4%
16	noun.motive	24	0.03%	2	0.04%	1	0.02%	0	0.00%	2	0.03%	8.3%
17	noun.object	2472	3.56%	249	5.49%	193	3.55%	96	3.74%	299	4.22%	12.1%
18	noun.person	5371	7.73%	416	9.17%	512	9.43%	168	6.54%	942	13.28%	17.5%
19	noun.phenomenon	444	0.64%	70	1.54%	55	1.01%	34	1.32%	75	1.06%	16.9%
20	noun.plant	7933	11.42%	67	1.48%	161	2.96%	19	0.74%	154	2.17%	1.9%
21	noun.possession	794	1.14%	61	1.34%	93	1.71%	48	1.87%	106	1.49%	13.4%
22	noun.process	579	0.83%	19	0.42%	61	1.12%	12	0.47%	27	0.38%	4.7%
23	noun.quantity	1159	1.67%	143	3.15%	50	0.92%	44	1.71%	69	0.97%	6.0%
24	noun.relation	525	0.76%	40	0.88%	45	0.83%	36	1.40%	82	1.16%	15.6%
25	noun.shape	383	0.55%	42	0.93%	24	0.44%	16	0.62%	44	0.62%	11.5%
26	noun.state	1857	2.67%	194	4.28%	220	4.05%	42	1.63%	209	2.95%	11.3%
27	noun.substance	2624	3.78%	284	6.26%	196	3.61%	113	4.40%	370	5.22%	14.1%
28	noun.time	810	1.17%	121	2.67%	72	1.33%	38	1.48%	107	1.51%	13.2%
	Total	69491		4538		5431		2569		7093		10.2%

<i>WN Lexicographer's file code</i>		WN		DE		FR		EE		CZ					
		Syns	% of wn	DE	%DE	FR	%FR	FR_W N%	EE	%EE	EE- WN%	CZ	%CZ	CZ- WN%	
29	verb.body	420	2.97%	88	4.11%	21.0%	59	2.11%	14.0%	16	3.32%	3.8%	19	1.42%	4.5%
30	verb.change	2827	19.97%	346	16.15%	12.2%	586	20.92%	20.7%	55	11.41%	1.9%	186	13.89%	6.6%
31	verb.cognition	771	5.45%	144	6.72%	18.7%	117	4.18%	15.2%	32	6.64%	4.2%	114	8.51%	14.8%
32	verb.communication	1445	10.21%	219	10.22%	15.2%	210	7.50%	14.5%	79	16.39%	5.5%	170	12.70%	11.8%
33	verb.competition	366	2.59%	48	2.24%	13.1%	27	0.96%	7.4%	3	0.62%	0.8%	30	2.24%	8.2%
34	verb.consumption	220	1.55%	50	2.33%	22.7%	60	2.14%	27.3%	11	2.28%	5.0%	22	1.64%	10.0%
35	verb.contact	1859	13.13%	228	10.64%	12.3%	279	9.96%	15.0%	26	5.39%	1.4%	104	7.77%	5.6%
36	verb.creation	757	5.35%	105	4.90%	13.9%	139	4.96%	18.4%	34	7.05%	4.5%	79	5.90%	10.4%
37	verb.emotion	268	1.89%	49	2.29%	18.3%	79	2.82%	29.5%	24	4.98%	9.0%	24	1.79%	9.0%
38	verb.motion	1750	12.36%	268	12.51%	15.3%	543	19.39%	31.0%	43	8.92%	2.5%	169	12.62%	9.7%
39	verb.perception	378	2.67%	88	4.11%	23.3%	96	3.43%	25.4%	20	4.15%	5.3%	31	2.32%	8.2%
40	verb.possession	814	5.75%	139	6.49%	17.1%	157	5.61%	19.3%	28	5.81%	3.4%	99	7.39%	12.2%
41	verb.social	1700	12.01%	275	12.83%	16.2%	279	9.96%	16.4%	90	18.67%	5.3%	198	14.79%	11.6%
42	verb.stative	532	3.76%	88	4.11%	16.5%	166	5.93%	31.2%	21	4.36%	3.9%	93	6.95%	17.5%
43	verb.weather	50	0.35%	8	0.37%	16.0%	4	0.14%	8.0%	0	0.00%	0.0%	1	0.07%	2.0%
	Total	14157		2143		15.1%	2801		19.8%	482		3.4%	1339		9.5%

IV. Conclusions:

Concrete Nouns:

- Global coverage of clusters is reasonable balanced across CZ and FR and across DE and EE. Methodology is different: CZ and FR start with translating and therefore have a larger coverage of translations, DE and EE start with constructing the Language Internal Relations and have a lower coverage of translations.
- Exceptional proportion of vehicles in Czech! No more work is needed here.
- Most clusters cover 7% of WN, which is the aim of Subset1
- Lower coverage for Animal, Group and Plant for all languages, Creature for EE, which is normal given the extreme proportion of these concepts in WN1.5

Abstract Nouns and Verbs

- Global coverage is balanced across CZ/FR and DE/EE
- Almost all clusters cover more than 7% of WN
- Lower coverage for Usage in FR (slightly for CZ and EE) and Existence in DE

Each of the wordnet builders can use the above data to improve the balancing of the different clusters, compared to the other wordnets.

V. Comparison of ILI-chains in EWN2

Hyperonym chains for Subset1 (DE, FR, EE)¹ converted to ILI-equivalences:

opstijgen (take off) stijgen (move to a higher position) verplaatsen (move location)
voortbewegen (move location) bewegen (move reflexive) bewegen (move intransitive)
veranderen (change)

00064108-v 01046072-v 01046072-v 01046072-v 01055491-v 01094615-v
00257753-v

Two kinds of measurements have been applied: sense-based (synset or ILI) and chain-based. The chain-based measurements have been divided into:

- node-coverage: synsets are covered but the relations may be different
- edge-coverage: synsets are covered but also the hyponymy relations between them

Language 1 chains: 1--2--3 & 4--5

Language 2 chains: 1--2 & 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. Both do have sub-chains: 1 sub-chain of length 3 in each language, and 2 sub-chains of length 2 (1 for each language).

There are two main objectives of the overall comparison: 1) to measure the degree of coverage and intersection of subset1 in EWN2 and 2) to be able to evaluate future improvements (if any) between EWN1 and EWN2. For this comparison we will refer to D014D015 deliverable (Vossen et al. 98). As in the case of subset1 the statistics have been extracted at three levels:

- 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 procedure for comparison has been the same that was performed for subset1 in EuroWordNet-1. For details of this procedure the reader can refer to (Vossen et al. 98). We will present here the main results and, when appropriate, comments about such results and the corresponding to the subset1 figures.

¹ For the Czech wordnet no ILI-chains have been received at the time of the comparison.

1. Evaluation of individual wordnets

	<i>ILI nodes</i>	<i>Tops</i>	<i>Leaves</i>	<i>Internal Nodes</i>	<i>EDGES</i>	<i>CHAINS</i>
WN15	60557	11	47110	13436	61123	53467
DE	4739	149	3777	960	5590	9271
FR	5523	603	3974	1518	4963	4432
EE	2742	104	1991	752	2809	2416
CZ	6754	N/A	N/A	N/A	N/A	N/A

	<i>ILI nodes</i>	<i>Tops</i>	<i>Leaves</i>	<i>Internal Nodes</i>	<i>EDGES</i>	<i>CHAINS</i>
WN15	11363	573	8446	2580	10816	8486
DE	2206	126	1794	478	3240	5058
FR	2534	124	2224	195	2417	2231
EE	535	72	352	139	494	372
CZ	1306	N/A	N/A	N/A	N/A	N/A

	<i>ILI nodes</i>	<i>Tops</i>	<i>Leaves</i>	<i>Internal Nodes</i>	<i>EDGES</i>	<i>CHAINS</i>
WN15	71920	584	55556	16016	71939	61953
DE	6945	275	5571	1438	8830	14329
FR	8057	727	6198	1713	7380	6663
EE	3277	176	2343	891	3303	2788
CZ	8060	N/A	N/A	N/A	N/A	N/A

A 1:1 ratio of ILI nodes and chains implies that the hierarchies are not tangled. This is the case for WN and FR. More chains than nodes, as is the case for DE, means a tangled hierarchy. This can be due to:

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

In the case of EE, we see that the number of chains is lower. This indicates a lack of hyperonyms, or translations. Furthermore, the number of tops is reasonable, except for FR nouns. It must be possible to improve this.

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

Tree-length

	<i>WN</i>		<i>DE</i>		<i>FR</i>		<i>SE</i>	
	<i>Freq.</i>	%	<i>Freq.</i>	%	<i>Freq.</i>	%	<i>Freq.</i>	%
1			62	0.51	572	12.91	49	2.00
2	33	0.06	206	1.71	54	1.22	87	3.55
3	522	0.97	744	6.18	241	5.44	235	9.60
4	2231	4.15	1883	15.63	551	12.43	408	16.66
5	5695	10.58	2197	18.24	869	19.61	534	21.80
6	12781	23.75	2079	17.26	895	20.19	396	16.17
7	11804	21.94	2101	17.44	719	16.22	301	12.29
8	8787	16.33	1359	11.28	307	6.93	297	12.13
9	6005	11.16	902	7.49	156	3.52	103	4.21
10	3358	6.24	365	3.03	47	1.06	38	1.55
11	1415	2.63	99	0.82	14	0.32	1	0.04
12	519	0.96	6	0.05	7	0.16		
13	367	0.68	14	0.12				
14	214	0.04	30	0.25				
15	75	0.14						
16	7	0.01						
Total	53813	100	12047	100	4432	100	2449	100
Average	7.19		6.04		5.17		5.47	

	<i>WN</i>		<i>GE</i>		<i>FR</i>		<i>ST</i>	
	<i>Freq.</i>	%	<i>Freq.</i>	%	<i>Freq.</i>	%	<i>Freq.</i>	%
1	236	2.78	30	0.32	9	0.40	15	3.70
2	1867	22.00	483	5.11	1101	49.35	103	25.43
3	2532	29.83	733	7.75	617	27.66	87	21.48
4	1959	23.08	1532	16.19	329	14.75	43	10.62
5	1028	12.11	1585	16.75	117	5.24	22	5.43
6	463	5.45	1352	14.29	46	2.06	25	6.17
7	250	2.95	992	10.49	11	0.49	23	5.68
8	109	1.28	1249	13.20	0	0.00	24	5.93
9	32	0.38	1019	10.77	1	0.04	28	6.91
10	10	0.12	382	4.04			20	4.94
11	2	0.02	88	0.93			8	1.98
12			14	0.15			6	1.48
13			2	0.02			1	0.25
Total	8488	100	9461	100	2231	100	405	100
Average	3.58		5.97		2.84		4.67	

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

<i>LENGTH</i>	<i>GE</i>	<i>FR</i>	<i>ST</i>	$\cap(GE,FR)$	$\cap(GE,ST)$	$\cap(FR,ST)$	$\cap(GE,FR,ST)$	<i>WN</i>
1	53453	53467	53456	53452	53348	53456	53348	53467
2	41464	53401	47070	41116	38677	47056	38637	53467
3	29832	52320	37641	29247	15978	37540	15959	53434
4	13446	44402	23739	12668	7302	23442	7255	52913
5	6749	34724	14013	6057	2320	13640	2299	50693
6	2477	22505	5856	2146	507	5552	494	45029
7	526	11720	2297	384	77	2001	71	32299
8	65	4673	463	38		401		20558
9	2	1290	28	1		16		11821
10		391	1			1		5881
11		237						2576
12		37						1176

Table 18 Coverage of partial noun chains of EDGES projected over WN1.5 structure

<i>LENGTH</i>	<i>GE</i>	<i>FR</i>	<i>ST</i>	$\cap(GE,FR)$	$\cap(GE,ST)$	$\cap(FR,ST)$	$\cap(GE,FR,ST)$	<i>WN</i>
1	30765	53401	42293	30153	20162	42240	20121	53467
2	11033	52320	30315	10439	2649	30183	2644	53434
3	1423	44402	18464	1333	511	18168	492	52913
4	434	34640	7874	229	3	7629		50693
5	112	22451	3214	77		3034		45029
6	3	11720	965	3		887		32299
7		4673	328			309		20558
8		1290						11821
9		391						5881
10		237						2576
11		37						1176

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

LENGTH	GE	FR	ST	$\cap(GE,FR)$	$\cap(GE,ST)$	$\cap(FR,ST)$	$\cap(GE,FR,ST)$	WN
1	7787	7241	6493	7121	6290	6385	6244	8486
2	4684	6430	2318	4356	1995	2277	1991	8250
3	2229	2944	636	1938	508	630	507	6383
4	904	1353	91	783	50	79	49	3853
5	370	527	25	306	19	19	19	1894
6	148	166		100				865
7	40	20						403
8	3	3						153
9	1	2						44

Table 20 Coverage of partial VERB chains of EDGES projected over WN1.5 structure

LENGTH	GE	FR	ST	$\cap(GE,FR)$	$\cap(GE,ST)$	$\cap(FR,ST)$	$\cap(GE,FR,ST)$	WN
1	2143	6430	1950	1992	337	1920	336	8250
2	256	2944	261	227		259		6383
3	68	1353	44	58		44		3853
4		527	19			19		1894
5		166						865
6		20						403
7		3						153
8		2						44

2. Conclusions:

- Average length of the chains is reasonably balanced, number of tops could be minimalized.
- Partial Node Coverage is compatible with EWN1 results and the Partial Edge Coverage exceeds EWN1 results despite the difference in coverage, which is very promising for the first subset. Partly explained by the closeness to WN structure.
- There is a significant difference between German and Estonian as compared to German/French and Estonian/French