The Restructured Core wordnets in EuroWordNet: Subset1

Version 3, Final
July 1, 1998

Contributors:
Piek Vossen, Laura Bloksma, University of Amsterdam
Salvador Climent, Maria Antonia Martí, Gabriel Oreggioni, University of Barcelona
Gerard Escudero, German Rigau, Horacio Rodriguez, Universidad Polytécnica de Catalunya
Adriana Roventini, Francesca Bertagna, Antonietta Alonge, Istituto di Linguistica del CNR, Pisa
Carol Peters, Istituto di Elaborazione della Informazione, CNR, Pisa
Wim Peters, University of Sheffield

Deliverable D014, D015, WP3, WP4
EuroWordNet, LE2-4003
<table>
<thead>
<tr>
<th>Identification number</th>
<th>LE-4003-D014-D015</th>
</tr>
</thead>
<tbody>
<tr>
<td>Type</td>
<td>Document and Lingware</td>
</tr>
<tr>
<td>Title</td>
<td>The Restructured Core wordnets in EuroWordNet: Subset1</td>
</tr>
<tr>
<td>Status</td>
<td>Final</td>
</tr>
<tr>
<td>Deliverable</td>
<td>D-014, D-015</td>
</tr>
<tr>
<td>Work Package</td>
<td>WP3 and WP4</td>
</tr>
<tr>
<td>Task</td>
<td>T3</td>
</tr>
<tr>
<td>Period covered</td>
<td>September 1997 – March 1998</td>
</tr>
<tr>
<td>Date</td>
<td>May 22, 1998</td>
</tr>
<tr>
<td>Version</td>
<td>2</td>
</tr>
<tr>
<td>Number of pages</td>
<td>84</td>
</tr>
</tbody>
</table>
| Authors               | • Piek Vossen, Laura Bloksma, University of Amsterdam  
• Salvador Climent, Toni Martin, Gabriel Torero Oreggioni University of Barcelona  
• Gerard Escudero, German Rigau, Horacio Rodriguez, Universidad Politecnica de Catalunya  
• Adriana Roventini, Francesca Bertagna, Antonietta Alonge, Istituto di Linguistica del CNR, Pisa  
• Carol Peters, Istituto di Elaborazione della Informazione, CNR, Pisa  
• Wim Peters, University of Sheffield |
| 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.let.uva.nl/~ewn. |
| Supplementary notes   | n.a. |
| Key words             | Linguistic Resources, Multilingual Wordnets, Language Engineering |
Abstract | This deliverable describes the First Subset for Nouns and Verbs in Dutch, Italian, Spanish and English. These First Subsets represent the cores of the wordnets: including the most important meanings on which the other meanings depend. The data are described in terms of tables that specify the synsets, entries, senses and relations, and by comparison with the top ontology distribution and the Parole lexicons. Furthermore, we have carried out two comparisons of the fragments. An in-depth comparison has been carried out for 18 semantic clusters, using the Polaris tool. An overall comparison has been carried out using a graph-matching toolkit developed by FUE. Finally, this deliverable describes the work done for updating the Inter-Lingual-Index (ILI) that interconnects the different wordnets. The conclusions of the overviews and comparisons are being used to guide the final building phase in EuroWordNet.

<table>
<thead>
<tr>
<th>Status of the abstract</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td>Received on</td>
<td></td>
</tr>
<tr>
<td>Recipient’s catalogue number</td>
<td></td>
</tr>
</tbody>
</table>
Executive Summary

This deliverable describes the First Subset for Nouns and Verbs in Dutch, Italian, Spanish and English. These First Subsets represent the cores of the wordnets: including the most important meanings on which the other meanings depend. The data for each language is described in terms of:

1. quantificational measure: tables that specify the synsets, entries, senses, the language internal relations and the equivalence relations.
2. the top ontology distribution of the synsets, indicating conceptual balancing of the subsets.
3. overlap with the Parole lexicons (as far as available).

Whereas the Spanish wordnet already has reached full coverage (advancing the planning), the Dutch wordnet has just covered the first subset with a higher density of language internal relations, and the Italian wordnet has full coverage but lacks equivalence relations. The distribution of the wordnets over the top-ontology was surprisingly balanced. Some slight imbalances for 1stOrder Entities have to be corrected. Similarly, the overlap with the top-frequent Parole entries is also very high. Missing entries can easily be added.

A better indication on the quality and compatibility is however given by comparing the consistency of the data across the wordnets. For this purpose we have carried out two comparisons of the fragments. An in-depth comparison has been carried out for 18 semantic clusters, using the Polaris tool. An overall comparison has been carried out using a graph-matching toolkit developed by FUE. Both comparisons showed promising results. The in-depth comparison of 18 fields showed reasonable intersections. Most of the mistakes are due to translation errors. Alternative classifications can be used to encode multiple hyperonym. A similar conclusion has been made from the overall comparison. There is a high degree of overlap between subsequences and sequences with 1 gap. By filling these gaps we can improve the coverage in a coordinated way. Furthermore, extremely tangled graphs (Dutch verbs) are mostly due to generation of wrong translations.

The following improvements will therefore be made to the wordnets in the next building phase:

- improve balancing of 1stOrderClusters (Dutch and Italian)
- extend with missing top-frequent Parole entries (Dutch, Italian, Spanish)
- extend coverage (Dutch)
- check translations of extremely long hyponymy chains (especially Dutch verbs)
- fill sequences with 1 gap (Italian, Spanish and Dutch)
- extend translations (Italian)
- improve translation heuristics (Spanish and Dutch)

Finally, this deliverable describes the work done for updating the Inter-Lingual-Index (ILI) that interconnects the different wordnets. The conclusions of the overviews and comparisons are being used to guide the final building phase in EuroWordNet.
## Table of Contents

1. General approach for building the wordnets........................................................................................................ 8
2. Overview results of Subset1 ....................................................................................................................................... 11
   2.1 Subset1 for the Dutch wordnet ......................................................................................................................... 11
   2.2 Subset1 for the Italian wordnet .......................................................................................................................... 14
   2.3 Subset1 for the Spanish wordnet ........................................................................................................................ 16
   2.4 Subset1 for the English wordnet ........................................................................................................................ 19
   2.5 Quantitative conclusions ....................................................................................................................................... 20
   2.6 Overlap with Parole lexicons ............................................................................................................................... 22
   2.7 Coverage of Subset1 over top concept clusters ................................................................................................. 24
3. Comparison of the first Subset .................................................................................................................................... 28
   3.1 Comparing specific semantic fields in the EuroWordNet database ..................................................................... 28
   3.2 Overall comparison of Subset1 ............................................................................................................................ 35
      3.2.1 Introduction .................................................................................................................................................. 35
      3.2.2 Evaluation of individual wordnets .............................................................................................................. 37
      3.2.3 Global evaluation .......................................................................................................................................... 40
4. Updating the ILI ........................................................................................................................................................... 48
   4.1 Clustering Methods .............................................................................................................................................. 49
      4.1.2 (Semi-)automatic clustering ......................................................................................................................... 50
         4.1.2.1 Sisters .................................................................................................................................................... 50
         4.1.2.2 Autohyponymy ...................................................................................................................................... 51
         4.1.2.3 Twins .................................................................................................................................................. 51
         4.1.2.4 Cousins .............................................................................................................................................. 51
         4.1.2.5 CoreLex .............................................................................................................................................. 52
      4.2 Testing automatically created sense groups .................................................................................................... 52
5 Conclusions .................................................................................................................................................................. 55

References ........................................................................................................................................................................ 56

Appendix I In-depth comparison of semantic clusters by different sites ........................................................................ 57
   Appendix Ia Comparing to the Dutch wordnet ......................................................................................................... 57
      Building................................................................................................................................................................ 58
      Comestibles ......................................................................................................................................................... 59
      Container ............................................................................................................................................................ 60
      Covering ............................................................................................................................................................ 61
      Feeling ................................................................................................................................................................. 62
      Phenomena ......................................................................................................................................................... 63
   Appendix Ib Comparing the Spanish wordnet ........................................................................................................ 64
      Garment ............................................................................................................................................................... 65
      Furniture ............................................................................................................................................................ 66
      Places ................................................................................................................................................................. 67
      Plants ................................................................................................................................................................. 68
      Sounds ............................................................................................................................................................... 69
      Cooking ............................................................................................................................................................. 70
   Appendix Ic Comparing the Italian wordnet ............................................................................................................ 71
      Animal ................................................................................................................................................................. 71
      Artist ................................................................................................................................................................. 72
      Worker .............................................................................................................................................................. 73
      Instrument ......................................................................................................................................................... 74
      Vehicle ............................................................................................................................................................... 75
      Movement ......................................................................................................................................................... 76
      Knowledge ......................................................................................................................................................... 77

Appendix II Software utilities for graph-comparison .................................................................................................... 78
**List of Tables**

<table>
<thead>
<tr>
<th>Table</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Table 1</td>
<td>First Subset Overview NL</td>
<td>11</td>
</tr>
<tr>
<td>Table 2</td>
<td>Language Internal Relations NL</td>
<td>12</td>
</tr>
<tr>
<td>Table 3</td>
<td>Equivalence Relations NL</td>
<td>13</td>
</tr>
<tr>
<td>Table 4</td>
<td>Status of the Language Internal Relations NL</td>
<td>13</td>
</tr>
<tr>
<td>Table 5</td>
<td>Status of the Equivalence Relations NL</td>
<td>13</td>
</tr>
<tr>
<td>Table 6</td>
<td>Reliability of the Equivalence Relations NL</td>
<td>13</td>
</tr>
<tr>
<td>Table 7</td>
<td>First Subset Overview IT</td>
<td>14</td>
</tr>
<tr>
<td>Table 8</td>
<td>Language Internal Relations IT</td>
<td>15</td>
</tr>
<tr>
<td>Table 9</td>
<td>Equivalence Relations IT</td>
<td>16</td>
</tr>
<tr>
<td>Table 10</td>
<td>First Subset Overview ES</td>
<td>16</td>
</tr>
<tr>
<td>Table 11</td>
<td>Language Internal Relations ES</td>
<td>17</td>
</tr>
<tr>
<td>Table 12</td>
<td>Equivalence Relations ES</td>
<td>18</td>
</tr>
<tr>
<td>Table 13</td>
<td>Reliability of Equivalence Relations ES</td>
<td>18</td>
</tr>
<tr>
<td>Table 14</td>
<td>First Subset Overview GB</td>
<td>19</td>
</tr>
<tr>
<td>Table 15</td>
<td>Language Internal Relations GB</td>
<td>20</td>
</tr>
<tr>
<td>Table 16</td>
<td>Equivalence Relations GB</td>
<td>20</td>
</tr>
<tr>
<td>Table 17</td>
<td>First Subset Overview: NL, ES, IT</td>
<td>21</td>
</tr>
<tr>
<td>Table 18</td>
<td>Overview of Language Internal Relations</td>
<td>21</td>
</tr>
<tr>
<td>Table 19</td>
<td>Overview of Equivalence Relations</td>
<td>22</td>
</tr>
<tr>
<td>Table 20</td>
<td>Coverage of Dutch Subset1 related to INL/Clex frequency</td>
<td>22</td>
</tr>
<tr>
<td>Table 21</td>
<td>Coverage of Spanish Subset1 of Parole Lexicons</td>
<td>23</td>
</tr>
<tr>
<td>Table 22</td>
<td>Coverage of WordNet1.5 compared to the English Parole Lexicon</td>
<td>23</td>
</tr>
<tr>
<td>Table 23</td>
<td>Synsets that are not clustered by the Top Ontology</td>
<td>24</td>
</tr>
<tr>
<td>Table 24</td>
<td>Nominal Synsets clustered as 1stOrder Concepts</td>
<td>24</td>
</tr>
<tr>
<td>Table 25</td>
<td>Nominal Synsets clustered as 2ndOrder Concepts</td>
<td>26</td>
</tr>
<tr>
<td>Table 26</td>
<td>Verbal Synsets clustered as 2ndOrder Concepts</td>
<td>27</td>
</tr>
<tr>
<td>Table 27</td>
<td>Nominal Synsets clustered as 3rdOrder Concepts</td>
<td>27</td>
</tr>
<tr>
<td>Table 28</td>
<td>Hyperonyms in the wordnets selected for the in-depth comparison</td>
<td>32</td>
</tr>
<tr>
<td>Table 29</td>
<td>Projections and Intersections of comparing the First Subset</td>
<td>33</td>
</tr>
<tr>
<td>Table 30</td>
<td>ILI chains for nouns</td>
<td>37</td>
</tr>
<tr>
<td>Table 31</td>
<td>ILI chains for verbs</td>
<td>37</td>
</tr>
<tr>
<td>Table 32</td>
<td>ILI chains for nouns and verbs</td>
<td>38</td>
</tr>
<tr>
<td>Table 33</td>
<td>ILI chains for nouns</td>
<td>38</td>
</tr>
<tr>
<td>Table 34</td>
<td>clean ILI chains for verbs</td>
<td>38</td>
</tr>
<tr>
<td>Table 35</td>
<td>Frequencies and ratios of noun chains / length /language</td>
<td>39</td>
</tr>
<tr>
<td>Table 36</td>
<td>Frequencies and ratios of verb chains / length /language</td>
<td>39</td>
</tr>
<tr>
<td>Table 37</td>
<td>Coverage of noun ILI records</td>
<td>40</td>
</tr>
<tr>
<td>Table 38</td>
<td>Coverage of verb ILI records</td>
<td>40</td>
</tr>
<tr>
<td>Table 39</td>
<td>Coverage of ILI records (total)</td>
<td>40</td>
</tr>
<tr>
<td>Table 40</td>
<td>Coverage of complete noun chains projected over WN1.5 structure</td>
<td>41</td>
</tr>
<tr>
<td>Table 41</td>
<td>Coverage of complete verb chains projected over WN1.5 structure</td>
<td>41</td>
</tr>
<tr>
<td>Table 42</td>
<td>Coverage of complete noun chains projected over Dutch wordnet</td>
<td>41</td>
</tr>
<tr>
<td>Table 43</td>
<td>Coverage of complete verb chains projected over Dutch wordnet</td>
<td>41</td>
</tr>
<tr>
<td>Table 44</td>
<td>Coverage of complete noun chains projected over Italian wordnet</td>
<td>41</td>
</tr>
<tr>
<td>Table 45</td>
<td>Coverage of complete verb chains projected over Italian wordnet</td>
<td>42</td>
</tr>
<tr>
<td>Table 46</td>
<td>Coverage of complete noun chains projected over Spanish wordnet</td>
<td>42</td>
</tr>
<tr>
<td>Table 47</td>
<td>Coverage of complete verb chains projected over Spanish wordnet</td>
<td>42</td>
</tr>
<tr>
<td>Table 48</td>
<td>Coverage of partial noun chains of NODES projected over WN1.5 structure</td>
<td>42</td>
</tr>
<tr>
<td>Table 49</td>
<td>Coverage of partial noun chains of EDGES projected over WN1.5 structure</td>
<td>43</td>
</tr>
<tr>
<td>Table 50</td>
<td>Coverage of partial VERB chains of NODES projected over WN1.5 structure</td>
<td>43</td>
</tr>
<tr>
<td>Table 51</td>
<td>Coverage of partial VERB chains of EDGES projected over WN1.5 structure</td>
<td>43</td>
</tr>
<tr>
<td>Table 52</td>
<td>Coverage of partial noun chains of NODES with 1 gap projected over WN1.5 structure</td>
<td>44</td>
</tr>
<tr>
<td>Table 53</td>
<td>Coverage of partial noun chains of EDGES with 1 gap projected over WN1.5 structure</td>
<td>44</td>
</tr>
<tr>
<td>Table 54</td>
<td>Coverage of partial VERB chains of NODES with 1 gap projected over WN1.5 structure</td>
<td>44</td>
</tr>
<tr>
<td>Table 55</td>
<td>Coverage of partial VERB chains of EDGES with 1 gap projected over WN1.5 structure</td>
<td>45</td>
</tr>
<tr>
<td>Table 56</td>
<td>Coverage of partial noun chains of NODES with 1 gap projected over Dutch wordnet</td>
<td>45</td>
</tr>
</tbody>
</table>
Table 57: Coverage of partial noun chains of EDGES with 1 gap projected over Dutch wordnet ........................................ 45
Table 58: Coverage of partial verb chains of NODES with 1 gap projected over Dutch wordnet ..................................... 45
Table 59: Coverage of partial verb chains of EDGES with 1 gap projected over Dutch wordnet ..................................... 45
Table 60: Coverage of partial noun chains of NODES with 1 gap projected over Italian wordnet ................................... 45
Table 61: Coverage of partial noun chains of EDGES with 1 gap projected over Italian wordnet ................................... 45
Table 62: Coverage of partial VERB chains of NODES with 1 gap projected over Italian wordnet ............................... 46
Table 63: Coverage of partial VERB chains of EDGES with 1 gap projected over Italian wordnet ............................... 46
Table 64: Coverage of partial noun chains of NODES with 1 gap projected over Spanish wordnet .............................. 46
Table 65: Coverage of partial noun chains of EDGES with 1 gap projected over Spanish wordnet .............................. 46
Table 66: Coverage of partial VERB chains of NODES with 1 gap projected over Spanish wordnet.......................... 46
Table 67: Coverage of partial VERB chains of EDGES with 1 gap projected over Spanish wordnet.......................... 46
Table 68: Coverage of partial noun chains of NODES with 2 gaps projected over WN1.5 structure ............................... 47
Table 69: Coverage of partial noun chains of EDGES with 2 gaps projected over WN1.5 structure ............................... 47
Table 70: Automatic derived generalizations and metonymy-relations ................................................................. 53
Table 71: Projection and Intersection increase Dutch-Spanish after adding sense-clusters to the ILI ............................ 54
Table 72: Errors generated by automatically derived Composite ILIs ........................................................................ 54
1. General approach for building the wordnets

The EuroWordNet database is being built (as much as possible) from available existing resources and databases with semantic information developed in various projects. In general, the wordnets are built in two major cycles as indicated by I and II in Figure 1 below. Each cycle consists of a building phase and a comparison phase:

1. Building a wordnet fragment
   1.1. Specification of an initial vocabulary
   1.2. Encoding of the language-internal relations
   1.3. Encoding of the equivalence relations
2. Comparing the wordnet fragments
   2.1. Loading of the wordnets in the EuroWordNet database
   2.2. Comparing and restructuring the fragments
   2.3. Measuring the overlap across the fragments

The building of a fragment is done using local tools and databases which are tailored to the specific nature and possibilities of the available resources. The available resources differ considerably in quality and explicitness of the data. Whereas some sites have the availability of partially structured networks between word senses, others start from genus words extracted from definitions that still have to be disambiguated in meaning.

After the specification of a fragment of the vocabulary, where each site uses similar criteria (there may again be differences due to the different starting points), globally, two approaches are followed for encoding the semantic relations:

**Merge model**: the selection is done in a local resource and the synsets and their language-internal relations are first developed separately, after which the equivalence relations are generated to WordNet1.5. This approach is followed for the Dutch and Italian wordnets.

**Expand model**: the selection is done in WordNet1.5 and the WordNet.1.5 synsets are translated (using bilingual dictionaries) into equivalent synsets in the other language. The wordnet relations are taken over and where necessary adapted to EuroWordNet. Possibly, monolingual resources are used to verify the wordnet relations imposed on non-English synsets. This approach is followed for the Spanish wordnet.

The Merge model results in a wordnet which is independent of WordNet1.5, possibly maintaining the language-specific properties. The Expand model will result in a wordnet which is very close to WordNet1.5 but which will also be biased by it. Whatever approach is followed also depends on the quality of the available resources.

After a production phase (step Ib and Ic in Figure 1) the results are converted to the EuroWordNet import format and loaded into the common database (step Ic). At that point various consistency checks are carried out, both formally and conceptually. By using the specific options in the EuroWordNet database it is then possible to further inspect and compare the data, to restructure relations where necessary and to measure the overlap in the fragments developed at the separate sites. Those meanings not covered by a site may be included in the extension of the vocabulary in the next building phase.
The overall design of the EuroWordNet database makes it possible to develop the individual language-specific wordnets relatively independently while guaranteeing a minimal level of compatibility. Nevertheless, some specific measures have been taken to enlarge the compatibility of the different resources:

1. The definition of a common set of so-called Base Concepts that is used as a starting point by all the sites to develop the cores of the wordnets. Base Concepts are meanings that play a major role in the wordnets: i.e. have many relations or high positions in the hierarchies.
2. The classification of the Base Concepts in terms of a Top Ontology.
3. The exchange of problems and possible solutions for encoding the relations for the Base Concepts.

The Base Concepts and the Top Ontology are further described in Deliverable D017D034D036 and in [Rodriguez et al. fc.]. In this document we describe the development of the first subset (Subset1) of wordnets in Dutch, Italian, Spanish and English, after the completion of one full cycle. Globally, the building has been carried out starting from the Base Concepts, extending top-down. The general criteria for Subset1 have been:

- All synsets linked to the common set of Base Concepts (1024 synsets).
- All relevant hyperonyms of the synsets related to the Base Concepts.
- The most important hyponyms (1 level down) of the synsets related to the Base Concept

In this way, Subset1 will at least include the core of the different wordnets, including the most important synsets on which more specific meanings depend. The cores will be developed mostly manually, whereas extensions will be derived using semi-automatic techniques.

---

1 The notion of Base Concepts should not be confused with Basic-Level Concepts as defined by Rosch (1977). Base Concepts are technically defined as the concepts with most relations. In most cases, they are more general than the Basic Level Concepts.
In addition, each site is free to add other concepts, suiting their local approach and starting point. These additions could be:

- synsets related via non-hyponymy relations (such meronymy, role/involvement, antonymy).
- synsets that are translatable to WordNet1.5 synsets.
- Easily extractable from the lexical resources that are available.
- Local Base Concepts, locally important concepts but still not part of the set of common Base Concepts.

The minimal set of synsets aimed at for Subset1 is 10,000 synsets, corresponding with about 20,000 word senses. For each of these synsets the following information has to be minimally specified:

- Hyperonym
- Synonyms (synset members)
- Equivalence relations to WordNet1.5

Optionally, any other relation could be added. The next figure gives an overview of the composition of the vocabulary.

In the case of SHE, a different approach has been followed. Because an English wordnet already exists, SHE has focussed on generating the relations which have been added in EuroWordNet with respect to WordNet1.5. Their selection has been based on the extractability of these relations. SHE has also restructured the ILI by adding missing glosses and grouping senses of words that show some kind of regular polysemy relation. This was necessary to provide a better matching across the wordnets.

The document is then structured in 3 main parts. In section 2, we give for each language overview tables of the covered subset, a comparison of the vocabulary with the Parole lexicons, and the coverage per Top Ontology clusters (e.g. Communication, Mental, Human, Animal). In section 3, we describe the results of comparing the semantic content of the subsets, by two methodologies: an in-depth comparison of a selection of semantic clusters and a global overall comparison of the complete subsets. Finally, the first results of restructuring the ILI are discussed in section 4. The final Subset1 can be obtained from ELRA (http://www.icp.inpg.fr/ELRA) or directly from the builders. Further information on the project and free samples of Subset1 can be downloaded from: http://www.hum.uva.nl/~ewn.

---

3 Parole is another EC project that builds lexicons for the most-frequent words with morpho-syntactic information.
2. Overview results of Subset1

The first subset is described in 3 ways for each site:

- number of entries, senses, and synsets covered and the number and kind of relations encoded: sections 2.1, 2.2, 2.3, 2.4, 2.5.
- comparison of the covered vocabulary with Parole lexicons: section 2.6.
- distribution of the vocabulary over the different top-ontology clusters: section 2.7.

2.1 Subset1 for the Dutch wordnet

AMS has followed the Merge Approach, where first a relatively stable Dutch wordnet has been built which has been linked (semi)-automatically to WordNet1.5. The building of the Dutch wordnet has been done by converting the usable relations from the Van Dale database (VLIS) to the EuroWordNet format. Next the converted relations have been verified manually (confirmed or deleted) and missing relations have been added. This manual coding has been done using a special editor, developed at AMS. Since initial hyponymy and synonymy relations were already present in the Van Dale database, we focussed on completing these relations and adding non-hyponymy relations.

The Dutch Subset1 has been based on:

- the common set of Base Concepts
- the local set of Base Concepts
- the hyponyms of the Base Concepts with more than 10 hyponyms themselves
- any other relation which has been manually added or confirmed

AMS has focussed on the encoding of a very solid and stable Subset1 with many different types of Language Internal Relations. We believe that creating a solid and rich semantic context will help both determining the more vague and fuzzy relations such as synonymy, and it will also help determining the equivalence relations with the ILI (WordNet1.5 synsets).

We have manually encoded the equivalence relations for 2,214 synsets. These include the equivalence relations to the common Base Concepts, and the equivalences to the local Base Concepts. All the other equivalences have been generated using a wordnet matching algorithm, partially based on the notion of Conceptual Density as developed by [Agirre and Rigau 1986]. This algorithm weights the senses of translations generated from a bilingual Dutch-English dictionary by comparing the distance of these senses to the senses of the translations of the Dutch semantic context of the word. The context is defined as all word senses that are directly related to it (by means of any semantic relation: hyperonyms, hyponyms, meronyms, etc.). The translation which best fits the translations of the context is selected. By translating synsets partly manually (creating very precise contexts) and by incrementally matching the translation, the best matching translations are generated, gradually improving the context. The current translations are generated after two incremental weightings, and the best 3 translations have been selected for Subset1. An advantage of taking the best 3 equivalence links is that there is a high reliability in coverage (see table below for a sample of Base Concepts). A drawback is that for each correct translation also 2 wrong translations may be generated. An evaluation of the method for more a larger part of the vocabulary will determine what is the best option. On the basis of the comparison it may turn out that we currently generate too much noise on addition to the correct translations.

Table 1: First Subset Overview NL

<table>
<thead>
<tr>
<th></th>
<th>Nouns</th>
<th>Verbs</th>
<th>Others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synsets</td>
<td>5917</td>
<td>3282</td>
<td>389</td>
<td>9588</td>
</tr>
<tr>
<td>Number of senses (variants)</td>
<td>10874</td>
<td>5915</td>
<td>1198</td>
<td>17987</td>
</tr>
<tr>
<td>X variants per synset</td>
<td>1.84</td>
<td>1.80</td>
<td>3.08</td>
<td>1.88</td>
</tr>
<tr>
<td>Corresponding to number of entries (words)</td>
<td>9555</td>
<td>4211</td>
<td>1070</td>
<td>14836</td>
</tr>
<tr>
<td>X senses per word</td>
<td>1.14</td>
<td>1.40</td>
<td>1.12</td>
<td>1.21</td>
</tr>
<tr>
<td>Language Internal Relations</td>
<td>16917</td>
<td>9486</td>
<td>432</td>
<td>26835</td>
</tr>
<tr>
<td>Average per synset</td>
<td>2.86</td>
<td>2.89</td>
<td>1.11</td>
<td>2.80</td>
</tr>
<tr>
<td>Equivalent Relations to ILI (WN1.5)</td>
<td>7664</td>
<td>6296</td>
<td>5</td>
<td>13965</td>
</tr>
<tr>
<td>Average per synset</td>
<td>1.30</td>
<td>1.92</td>
<td>0.01</td>
<td>1.46</td>
</tr>
<tr>
<td>Synset without ILI</td>
<td>1578</td>
<td>394</td>
<td>385</td>
<td>2357</td>
</tr>
</tbody>
</table>
Table 2: Language Internal Relations NL

<table>
<thead>
<tr>
<th>Language Internal Relations</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synsets</td>
<td>5917</td>
<td>3282</td>
<td>389</td>
<td>9588</td>
</tr>
<tr>
<td>BE_IN_STATE</td>
<td>93</td>
<td></td>
<td></td>
<td>93</td>
</tr>
<tr>
<td>CAUSES</td>
<td>140</td>
<td>609</td>
<td></td>
<td>749</td>
</tr>
<tr>
<td>HAS_HYPERONYM</td>
<td>6169</td>
<td>3588</td>
<td></td>
<td>9757</td>
</tr>
<tr>
<td>HAS_HYPONYM</td>
<td>6169</td>
<td>3588</td>
<td></td>
<td>9757</td>
</tr>
<tr>
<td>HAS_HOLONYM</td>
<td>275</td>
<td></td>
<td></td>
<td>275</td>
</tr>
<tr>
<td>HAS_HOLO_LOCATION</td>
<td>84</td>
<td></td>
<td></td>
<td>84</td>
</tr>
<tr>
<td>HAS_HOLO_MADEOF</td>
<td>97</td>
<td></td>
<td></td>
<td>97</td>
</tr>
<tr>
<td>HAS_HOLO_MEMBER</td>
<td>108</td>
<td></td>
<td></td>
<td>108</td>
</tr>
<tr>
<td>HAS_HOLO_PART</td>
<td>444</td>
<td></td>
<td></td>
<td>444</td>
</tr>
<tr>
<td>HAS_HOLO_PORTION</td>
<td>66</td>
<td></td>
<td></td>
<td>66</td>
</tr>
<tr>
<td>HAS_MERONYM</td>
<td>286</td>
<td></td>
<td></td>
<td>286</td>
</tr>
<tr>
<td>HAS_MERO_LOCATION</td>
<td>84</td>
<td></td>
<td></td>
<td>84</td>
</tr>
<tr>
<td>HAS_MERO_MADEOF</td>
<td>97</td>
<td></td>
<td></td>
<td>97</td>
</tr>
<tr>
<td>HAS_MERO_MEMBER</td>
<td>110</td>
<td></td>
<td></td>
<td>110</td>
</tr>
<tr>
<td>HAS_MERO_PART</td>
<td>442</td>
<td></td>
<td></td>
<td>442</td>
</tr>
<tr>
<td>HAS_MERO_PORTION</td>
<td>65</td>
<td></td>
<td></td>
<td>65</td>
</tr>
<tr>
<td>HAS_SUBEVENT</td>
<td>99</td>
<td>109</td>
<td></td>
<td>208</td>
</tr>
<tr>
<td>HAS_XPOS_HYPERONYM</td>
<td>9</td>
<td>34</td>
<td>5</td>
<td>48</td>
</tr>
<tr>
<td>HAS_XPOS_HYPONYM</td>
<td>34</td>
<td>13</td>
<td>1</td>
<td>48</td>
</tr>
<tr>
<td>INVOLVED</td>
<td>54</td>
<td>81</td>
<td></td>
<td>135</td>
</tr>
<tr>
<td>INVOLVED_AGENT</td>
<td>4</td>
<td>29</td>
<td></td>
<td>33</td>
</tr>
<tr>
<td>INVOLVED_DIRECTION</td>
<td>28</td>
<td>1</td>
<td></td>
<td>29</td>
</tr>
<tr>
<td>INVOLVED_INSTRUMENT</td>
<td>5</td>
<td>232</td>
<td></td>
<td>237</td>
</tr>
<tr>
<td>INVOLVED_LOCATION</td>
<td>195</td>
<td>21</td>
<td></td>
<td>216</td>
</tr>
<tr>
<td>INVOLVED_PATIENT</td>
<td>16</td>
<td>285</td>
<td></td>
<td>301</td>
</tr>
<tr>
<td>INVOLVED_SOURCE_DIRECTION</td>
<td>2</td>
<td>1</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>INVOLVED_TARGET_DIRECTION</td>
<td>215</td>
<td>20</td>
<td></td>
<td>235</td>
</tr>
<tr>
<td>IS_CAUSED_BY</td>
<td>81</td>
<td>208</td>
<td>320</td>
<td>609</td>
</tr>
<tr>
<td>IS_SUBEVENT_OF</td>
<td>91</td>
<td>128</td>
<td></td>
<td>219</td>
</tr>
<tr>
<td>NEAR_ANTONYM</td>
<td>132</td>
<td>217</td>
<td></td>
<td>349</td>
</tr>
<tr>
<td>NEAR_SYNONYM</td>
<td>138</td>
<td>81</td>
<td></td>
<td>219</td>
</tr>
<tr>
<td>ROLE</td>
<td>32</td>
<td></td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>ROLE_AGENT</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>ROLE_DIRECTION</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>ROLE_INSTRUMENT</td>
<td>259</td>
<td></td>
<td></td>
<td>259</td>
</tr>
<tr>
<td>ROLE_LOCATION</td>
<td>26</td>
<td></td>
<td></td>
<td>26</td>
</tr>
<tr>
<td>ROLE_PATIENT</td>
<td>482</td>
<td></td>
<td></td>
<td>482</td>
</tr>
<tr>
<td>ROLE_SOURCE_DIRECTION</td>
<td>16</td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>ROLE_TARGET_DIRECTION</td>
<td>18</td>
<td></td>
<td></td>
<td>18</td>
</tr>
<tr>
<td>STATE_OF</td>
<td>9</td>
<td>6</td>
<td>79</td>
<td>94</td>
</tr>
<tr>
<td>XPOS_NEAR_ANTONYM</td>
<td>3</td>
<td>3</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>XPOS_NEAR_SYNONYM</td>
<td>237</td>
<td>232</td>
<td>27</td>
<td>496</td>
</tr>
<tr>
<td>Total</td>
<td>16917</td>
<td>9486</td>
<td>432</td>
<td>26835</td>
</tr>
<tr>
<td>Average per synset</td>
<td>2.86</td>
<td>2.89</td>
<td>1.11</td>
<td>2.80</td>
</tr>
</tbody>
</table>
Table 3: Equivalence Relations NL

<table>
<thead>
<tr>
<th>Equivalence Relations</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ_NEAR_SYNONYM</td>
<td>6025</td>
<td>5883</td>
<td>11908</td>
</tr>
<tr>
<td>EQ_SYNONYM</td>
<td>1370</td>
<td>375</td>
<td>1745</td>
</tr>
<tr>
<td>EQ_HAS_HYPERONYM</td>
<td>174</td>
<td>22</td>
<td>196</td>
</tr>
<tr>
<td>EQ_HAS_HYPONYM</td>
<td>85</td>
<td>11</td>
<td>96</td>
</tr>
<tr>
<td>EQ_INVOLVED</td>
<td>4</td>
<td>4</td>
<td></td>
</tr>
<tr>
<td>EQ_IS_CAUSED_BY</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>EQ_HAS_HOLONYM</td>
<td>3</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>EQ_HAS_MERONYM</td>
<td>6</td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>7664</td>
<td>6296</td>
<td>13960</td>
</tr>
</tbody>
</table>

The next table indicates the number of relations taken over from the Van Dale database or added manually:

Table 4: Status of the Language Internal Relations NL

<table>
<thead>
<tr>
<th>Language Internal Relations</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Other</th>
<th>Total</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Vlis &amp; Okay</td>
<td>2867</td>
<td>2025</td>
<td>3</td>
<td>5075</td>
<td>64.16%</td>
</tr>
<tr>
<td>Vlis &amp; ?</td>
<td>2317</td>
<td>517</td>
<td>1</td>
<td>2835</td>
<td>35.84%</td>
</tr>
<tr>
<td>Vlis Total</td>
<td>5184</td>
<td>2722</td>
<td>4</td>
<td>7910</td>
<td>60.40%</td>
</tr>
<tr>
<td>Manual &amp; Okay</td>
<td>437</td>
<td>2915</td>
<td>286</td>
<td>3638</td>
<td>70.14%</td>
</tr>
<tr>
<td>Manual &amp; ?</td>
<td>1238</td>
<td>258</td>
<td>53</td>
<td>1549</td>
<td>29.86%</td>
</tr>
<tr>
<td>Manual Total</td>
<td>1675</td>
<td>3173</td>
<td>339</td>
<td>5187</td>
<td>39.60%</td>
</tr>
<tr>
<td>Total</td>
<td>6859</td>
<td>5895</td>
<td>343</td>
<td>13097</td>
<td></td>
</tr>
</tbody>
</table>

The next table gives the distribution of the manually generated translations and the translations generated by the matching heuristics:

Table 5: Status of the Equivalence Relations NL

<table>
<thead>
<tr>
<th>Language External relations</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Other</th>
<th>Total</th>
<th>Percentages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Heuristics &amp; Okay</td>
<td>303</td>
<td>69</td>
<td>372</td>
<td></td>
<td>3.17%</td>
</tr>
<tr>
<td>Heuristics &amp; ?</td>
<td>5619</td>
<td>5760</td>
<td>11379</td>
<td></td>
<td>96.83%</td>
</tr>
<tr>
<td>Heuristics Total</td>
<td>5922</td>
<td>5829</td>
<td>0</td>
<td>11751</td>
<td>84.15%</td>
</tr>
<tr>
<td>Manual &amp; Okay</td>
<td>691</td>
<td>151</td>
<td>5</td>
<td>847</td>
<td>38.26%</td>
</tr>
<tr>
<td>Manual &amp; ?</td>
<td>1051</td>
<td>316</td>
<td>1367</td>
<td></td>
<td>61.74%</td>
</tr>
<tr>
<td>Manual Total</td>
<td>1742</td>
<td>467</td>
<td>5</td>
<td>2214</td>
<td>15.85%</td>
</tr>
<tr>
<td>Total</td>
<td>7664</td>
<td>6296</td>
<td>5</td>
<td>13965</td>
<td></td>
</tr>
</tbody>
</table>

For a sample of Base Concepts we generated the equivalence relations by heuristics and checked the quality of the scores. The next tables gives the reliability of taking the top-3 equivalence relations generated by the heuristics. The table is differentiated for a sample of 1stOrderEntities (FOEs) and 2nd/3rdOrderEntities (HOEs):

Table 6: Reliability of the Equivalence Relations NL

<table>
<thead>
<tr>
<th>Matching Rank</th>
<th>HOEs</th>
<th>FOEs</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>No of synsets</td>
<td>Perc.</td>
</tr>
<tr>
<td>1st score</td>
<td>49</td>
<td>44.95%</td>
</tr>
<tr>
<td>2nd score</td>
<td>36</td>
<td>33.03%</td>
</tr>
<tr>
<td>3rd score</td>
<td>9</td>
<td>8.26%</td>
</tr>
<tr>
<td>&gt;</td>
<td>15</td>
<td>13.76%</td>
</tr>
<tr>
<td>Sum</td>
<td>109</td>
<td></td>
</tr>
</tbody>
</table>

The table shows that in about 50% the 1st score is also the correct translation, and in about 82-87% the correct one is among the top-3. Note that these BCs are the most difficult cases to translate. For more specific concepts the rates will go up.
2.2 Subset1 for the Italian wordnet

At Pisa we have first automatically extracted first level hyponyms of the common Base Concepts (CBCs) from our LDB (which contains data from different sources). As far as the nouns are concerned a sense disambiguation of hyponyms had already been performed within other research projects, thus we only had to revise taxonomies in order to see if hyponyms had been properly assigned to each taxonomy. With respect to the verbs, instead, disambiguated taxonomies had been previously built only for some of our BCs. Thus we had to manually perform a sense disambiguation of most of the taxonomies built. Then, since with respect to other relations to be encoded in EWN our database contained only some information already partially encoded in previous projects (e.g. Acquilex, Delis: synonymy, part-of, set-of, deverbal, deadjectival for nouns; synonymy for verbs), we had to manually add all the other relations, by analysing mainly definitions, but also other information available (e.g., examples provided for each word sense in our source). The PSA Subset1 has been based on:

- the common set of Base Concepts
- the local set of Base Concepts
- all first level hyponyms of the Base Concepts
- for some taxonomies, also other level hyponyms
- any other relation which has been manually added, by analysing mainly definitions

By using a semi-automatic procedure, part of the data elaborated has been already mapped to WN 1.5., but this work is still in progress. For about 3,091 synsets we have semi-automatically generated the equivalence relations, which have however been verified manually.

Table 7: First Subset Overview IT

<table>
<thead>
<tr>
<th></th>
<th>Nouns</th>
<th>Verbs</th>
<th>Others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synsets</td>
<td>18934</td>
<td>3692</td>
<td>1581</td>
<td>24207</td>
</tr>
<tr>
<td>Number of senses (variants)</td>
<td>19646</td>
<td>4577</td>
<td>1587</td>
<td>25810</td>
</tr>
<tr>
<td>X variants per synset</td>
<td>1.03</td>
<td>1.24</td>
<td>1</td>
<td>1.09</td>
</tr>
<tr>
<td>Corresponding to number of entries (words)</td>
<td>13965</td>
<td>3170</td>
<td>1</td>
<td>17135</td>
</tr>
<tr>
<td>X senses per word</td>
<td>1.40</td>
<td>1.44</td>
<td>1</td>
<td>1.50</td>
</tr>
<tr>
<td>Language Internal Relations</td>
<td>47090</td>
<td>9070</td>
<td>1</td>
<td>56160</td>
</tr>
<tr>
<td>Average per synset</td>
<td>2.48</td>
<td>2.45</td>
<td>2</td>
<td>2.32</td>
</tr>
<tr>
<td>Equivalent Relations to ILI (WN1.5)</td>
<td>5124</td>
<td>653</td>
<td>3</td>
<td>5777</td>
</tr>
<tr>
<td>Average per synset</td>
<td>0.27</td>
<td>0.17</td>
<td>0</td>
<td>0.22</td>
</tr>
<tr>
<td>Synset without ILI</td>
<td>13957</td>
<td>3109</td>
<td>1581</td>
<td>18647</td>
</tr>
</tbody>
</table>
Table 8: Language Internal Relations IT

<table>
<thead>
<tr>
<th>Language Internal Relations</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synsets</td>
<td>18934</td>
<td>3692</td>
<td>1581</td>
<td>24207</td>
</tr>
<tr>
<td>BE_IN_STATE</td>
<td>123</td>
<td></td>
<td></td>
<td>123</td>
</tr>
<tr>
<td>CAUSES</td>
<td></td>
<td>569</td>
<td></td>
<td>569</td>
</tr>
<tr>
<td>HAS_HYPERONYM</td>
<td>18654</td>
<td>3651</td>
<td></td>
<td>22305</td>
</tr>
<tr>
<td>HAS_HYPONYM</td>
<td>18654</td>
<td>3651</td>
<td></td>
<td>22305</td>
</tr>
<tr>
<td>HAS_HOLONYM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAS_HOLO_LOCATION</td>
<td>6</td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>HAS_HOLO_MADEOF</td>
<td>1</td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>HAS_HOLO_MEMBER</td>
<td>34</td>
<td></td>
<td></td>
<td>34</td>
</tr>
<tr>
<td>HAS_HOLO_PART</td>
<td>290</td>
<td></td>
<td></td>
<td>290</td>
</tr>
<tr>
<td>HAS_HOLO_PORTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAS_MERONYM</td>
<td>264</td>
<td></td>
<td></td>
<td>264</td>
</tr>
<tr>
<td>HAS_MERO_LOCATION</td>
<td>5</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>HAS_MERO_MADEOF</td>
<td>165</td>
<td></td>
<td></td>
<td>165</td>
</tr>
<tr>
<td>HAS_MERO_MEMBER</td>
<td>186</td>
<td></td>
<td></td>
<td>186</td>
</tr>
<tr>
<td>HAS_MERO_PART</td>
<td>219</td>
<td></td>
<td></td>
<td>219</td>
</tr>
<tr>
<td>HAS_MERO_PORTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>HAS_SUBEVENT</td>
<td>82</td>
<td></td>
<td></td>
<td>82</td>
</tr>
<tr>
<td>HAS_XPOS_HYPERONYM</td>
<td>2</td>
<td></td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>HAS_XPOS_HYPONYM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INVOLVED</td>
<td>755</td>
<td>36</td>
<td>10</td>
<td>755</td>
</tr>
<tr>
<td>INVOLVED_AGENT</td>
<td>94</td>
<td>5</td>
<td>10</td>
<td>94</td>
</tr>
<tr>
<td>INVOLVED_LOCATION</td>
<td>1</td>
<td>5</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td>INVOLVED_INSTRUMENT</td>
<td>101</td>
<td>5</td>
<td>101</td>
<td>101</td>
</tr>
<tr>
<td>INVOLVED_SOURCE_DIRECTION</td>
<td>53</td>
<td>5</td>
<td>53</td>
<td>53</td>
</tr>
<tr>
<td>INVOLVED_TARGET_DIRECTION</td>
<td>18</td>
<td>5</td>
<td>18</td>
<td>18</td>
</tr>
<tr>
<td>IS_CAUSED_BY</td>
<td>32</td>
<td></td>
<td></td>
<td>32</td>
</tr>
<tr>
<td>IS_SUBEVENT_OF</td>
<td>5</td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>NEAR_ANTONYM (ANTONYM)</td>
<td>20</td>
<td>4</td>
<td>24</td>
<td></td>
</tr>
<tr>
<td>NEAR_SYNONYM</td>
<td>221</td>
<td>4</td>
<td>225</td>
<td></td>
</tr>
<tr>
<td>ROLE</td>
<td>21</td>
<td></td>
<td></td>
<td>21</td>
</tr>
<tr>
<td>ROLE_AGENT</td>
<td>1095</td>
<td>4</td>
<td>1099</td>
<td>1099</td>
</tr>
<tr>
<td>ROLE_DIRECTION</td>
<td></td>
<td>80</td>
<td></td>
<td>80</td>
</tr>
<tr>
<td>ROLE_INSTRUMENT</td>
<td></td>
<td>80</td>
<td></td>
<td>80</td>
</tr>
<tr>
<td>ROLE_LOCATION</td>
<td>51</td>
<td></td>
<td></td>
<td>51</td>
</tr>
<tr>
<td>ROLE_PATIENT</td>
<td>16</td>
<td></td>
<td></td>
<td>16</td>
</tr>
<tr>
<td>ROLE_SOURCE_DIRECTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ROLE_TARGET_DIRECTION</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STATE_OF</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XPOS_NEAR_ANTONYM</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>XPOS_NEAR_SYNONYM</td>
<td>7505</td>
<td>7505</td>
<td></td>
<td>7505</td>
</tr>
<tr>
<td>Total</td>
<td>47090</td>
<td>9070</td>
<td>56160</td>
<td></td>
</tr>
<tr>
<td>Average per synset</td>
<td>2.48</td>
<td>2.45</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 9: Equivalence Relations IT

<table>
<thead>
<tr>
<th>Equivalence Relations</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ_SYNONYM</td>
<td>3697</td>
<td>307</td>
<td>4004</td>
</tr>
<tr>
<td>EQ_NEAR_SYNONYM</td>
<td>631</td>
<td>259</td>
<td>890</td>
</tr>
<tr>
<td>EQ_HAS_HYPERONYM</td>
<td>1947</td>
<td>77</td>
<td>2024</td>
</tr>
<tr>
<td>EQ_HAS_HYPONYM</td>
<td>10</td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>Total</td>
<td>6275</td>
<td>653</td>
<td>6928</td>
</tr>
</tbody>
</table>

2.3 Subset1 for the Spanish wordnet

FUE has followed the Expand Approach, differentiated for Nouns and Verbs. This approach is based on automatic assignment of Spanish Words to WordNet 1.5 synsets plus further manual revision of the Spanish synsets and relations thus built. The main consequences of this approach are the following:

- A great number of Spanish synsets have been built thus almost reaching already the quantity expected for the end of the project
- The Spanish WordNet (SWN) semantic network is already achieved in terms of the main relations importable from WordNet 1.5 (hypernymy/hyponymy for all Nouns and Verbs, meronymy for Nouns, and causation for Verbs)
- All synsets in the Spanish SWN have an equivalence link to the ILI
- The building of the SWN in this phase has been constrained by conditions of translability to English
- Spanish Nouns automatically assigned to synsets are subject to a degree of confidence, which in any case ranks at least above 85%; all Verbs have been manually checked and corrected; all relations in Subset1 either have been manually built or have a confidence score of 100%.

The building of the SWN Subset1 has proceeded as follows. For Verbs, the PIRAPIDES database (developed by the Universities of Barcelona and Maryland in a join project, see [Dorr et al 1997] ) has been used. It consists of 3600 English verb forms organized around Levin's Semantic Classes [Levin 1993], connected to WN1.5 senses, and ambiguously translated to Spanish. It also contains thematic role and diathesis information. Using the latter information and other linguistic knowledge, the database has been manually processed to produce correct SWN synsets. Subset1 includes those which have been already processed; the rest will be included in Subset2. For Nouns, a methodology to map Spanish word forms to WN1.5 synsets using bilingual dictionaries (described in [Atserias et al. 1997]) has been followed. By this procedure, several heuristics have been manually tested using a local lexicological environment in order to choose those which give higher mapping confidence ratios, thus building the appropriate SWN synsets. Furthermore, a number of synsets, including the common set of Base Concepts and the Spanish counterparts of the higher levels in the WN1.5 taxonomy have been manually built. Relations between synsets have been manually checked to include those which are importable from WN1.5 to the SWN. Those which are not will be included in Subset2. Quantity and Quality of the SWN Subset1 can be seen in the tables below.

Table 10: First Subset Overview ES

<table>
<thead>
<tr>
<th></th>
<th>Nouns</th>
<th>Verbs</th>
<th>Others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synsets</td>
<td>18577</td>
<td>2602</td>
<td>0</td>
<td>21179</td>
</tr>
<tr>
<td>number of senses (variants)</td>
<td>41292</td>
<td>6795</td>
<td>0</td>
<td>48087</td>
</tr>
<tr>
<td>X variants per synset</td>
<td>2.22</td>
<td>2.61</td>
<td>0</td>
<td>2.27</td>
</tr>
<tr>
<td>Corresponding to number of entries (words)</td>
<td>23216</td>
<td>2278</td>
<td>0</td>
<td>25494</td>
</tr>
<tr>
<td>X senses per word</td>
<td>1.77</td>
<td>2.98</td>
<td>0</td>
<td>1.88</td>
</tr>
<tr>
<td>Language Internal Relations</td>
<td>40559</td>
<td>3749</td>
<td>0</td>
<td>44308</td>
</tr>
<tr>
<td>Average per synset</td>
<td>2.18</td>
<td>1.44</td>
<td>0</td>
<td>2.09</td>
</tr>
<tr>
<td>Equivalent Relations to ILI (WN1.5)</td>
<td>18634</td>
<td>2602</td>
<td>0</td>
<td>21236</td>
</tr>
<tr>
<td>Average per synset</td>
<td>1.00</td>
<td>1.00</td>
<td>0</td>
<td>1.00</td>
</tr>
<tr>
<td>Synset without ILI</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Percentage of Synsets without translation</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
</tbody>
</table>
## Table 11: Language Internal Relations ES

<table>
<thead>
<tr>
<th>Language Internal Relations</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>BE_IN_STATE</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>CAUSES</td>
<td>0</td>
<td>40</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>HAS_HYPERONYM</td>
<td>18907</td>
<td>1830</td>
<td>0</td>
<td>20737</td>
</tr>
<tr>
<td>HAS_HYPONYM</td>
<td>18907</td>
<td>1830</td>
<td>0</td>
<td>20737</td>
</tr>
<tr>
<td>HAS_HOLONYM</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HAS_HOLO_LOCATION</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HAS_HOLO_MADEOF</td>
<td>75</td>
<td>0</td>
<td>0</td>
<td>75</td>
</tr>
<tr>
<td>HAS_HOLO_MEMBER</td>
<td>188</td>
<td>0</td>
<td>0</td>
<td>188</td>
</tr>
<tr>
<td>HAS_HOLO_PART</td>
<td>1103</td>
<td>0</td>
<td>0</td>
<td>1103</td>
</tr>
<tr>
<td>HAS_HOLO_PORTION</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HAS_MERONYM</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HAS_MERO_LOCATION</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HAS_MERO_MADEOF</td>
<td>75</td>
<td>0</td>
<td>0</td>
<td>75</td>
</tr>
<tr>
<td>HAS_MERO_MEMBER</td>
<td>188</td>
<td>0</td>
<td>0</td>
<td>188</td>
</tr>
<tr>
<td>HAS_MERO_PART</td>
<td>1103</td>
<td>0</td>
<td>0</td>
<td>1103</td>
</tr>
<tr>
<td>HAS_MERO_PORTION</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HAS_SUBEVENT</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>HAS_XPOS_HYPERONYM</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>HAS_XPOS_HYPONYM</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INVOLVED</td>
<td>0</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>INVOLVED_AGENT</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>INVOLVED_DIRECTION</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INVOLVED_INSTRUMENT</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>INVOLVED_LOCATION</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INVOLVED_PATIENT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INVOLVED_SOURCE_DIRECTION</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>INVOLVED_TARGET_DIRECTION</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IS_CAUSED_BY</td>
<td>0</td>
<td>40</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>IS_SUBEVENT_OF</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>NEAR_ANTONYM</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NEAR_SYNONYM</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>6</td>
</tr>
<tr>
<td>ROLE</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td>ROLE_AGENT</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>ROLE_DIRECTION</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ROLE_INSTRUMENT</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>ROLE_LOCATION</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ROLE_PATIENT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ROLE_SOURCE_DIRECTION</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ROLE_TARGET_DIRECTION</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>STATE_OF</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>XPOS_NEAR_ANTONYM</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>XPOS_NEAR_SYNONYM</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>40559</td>
<td>3749</td>
<td>0</td>
<td>44308</td>
</tr>
</tbody>
</table>
Table 12: Equivalence Relations ES

<table>
<thead>
<tr>
<th>Equivalence Relations</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ_NEAR_SYNONYM</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EQ_SYNONYM</td>
<td>18577</td>
<td>2602</td>
<td>21179</td>
</tr>
<tr>
<td>EQ_HAS_HYPERONYM</td>
<td>40</td>
<td>0</td>
<td>40</td>
</tr>
<tr>
<td>EQ_HAS_HYPONYM</td>
<td>14</td>
<td>0</td>
<td>14</td>
</tr>
<tr>
<td>EQ_INVOLVED</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EQ_IS_CAUSED_BY</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>EQ_HAS_HOLONYM</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>EQ_HAS_MERONYM</td>
<td>2</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>18634</td>
<td>2602</td>
<td>21236</td>
</tr>
</tbody>
</table>

The next table indicates the reliability of generated translation:

Table 13: Reliability of Equivalence Relations ES

<table>
<thead>
<tr>
<th>Confidence (Variants)</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>100% (Manual)</td>
<td>5041</td>
<td>6795</td>
<td>11836</td>
</tr>
<tr>
<td>&gt;97%</td>
<td>403</td>
<td>0</td>
<td>403</td>
</tr>
<tr>
<td>&gt;95%</td>
<td>304</td>
<td>0</td>
<td>304</td>
</tr>
<tr>
<td>&gt;93%</td>
<td>1598</td>
<td>0</td>
<td>1598</td>
</tr>
<tr>
<td>&gt;86%</td>
<td>27649</td>
<td>0</td>
<td>27649</td>
</tr>
<tr>
<td>&gt;85%</td>
<td>4625</td>
<td>0</td>
<td>4625</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>39620</td>
<td>6795</td>
<td>46415</td>
</tr>
</tbody>
</table>
2.4 Subset1 for the English wordnet

Sheffield has concentrated on morphological derivational relations between nouns and verbs in order to create cross-part-of-speech relations expressing morphological as well as semantic links. This subset contains morphological derivational relations between nouns and verbs where the verb has been the base form for the derivational process. The data has been obtained from the CELEX database, in which suffixation and conversion (zero-derivation) processes have been identified and a hierarchical morphological decomposition of the derived form has been performed. CELEX noun-verb pairs with a derivational relation have then been matched against the WordNet wordforms. The WordNet senses of the selected pairs have been manually compared and semantic relations have been manually assigned. The English subset has been based on an extended base concept set of 2277 noun synsets and 567 verb synsets.

Table 14: First Subset Overview GB

<table>
<thead>
<tr>
<th></th>
<th>Nouns</th>
<th>Verbs</th>
<th>Others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synsets</td>
<td>968</td>
<td>894</td>
<td></td>
<td>1862</td>
</tr>
<tr>
<td>Number of senses (variants)</td>
<td>2235</td>
<td>3035</td>
<td></td>
<td>5270</td>
</tr>
<tr>
<td>X variants per synset</td>
<td>1.99</td>
<td>2.69</td>
<td></td>
<td>2.34</td>
</tr>
<tr>
<td>Corresponding to number of entries (words)</td>
<td>1927</td>
<td>2411</td>
<td></td>
<td>4338</td>
</tr>
<tr>
<td>X senses per word</td>
<td>1.16</td>
<td>1.26</td>
<td></td>
<td>1.21</td>
</tr>
<tr>
<td>Language Internal Relations</td>
<td>2785</td>
<td>2616</td>
<td></td>
<td>5401</td>
</tr>
<tr>
<td>Average per synset</td>
<td>2.88</td>
<td>2.92</td>
<td></td>
<td>2.9</td>
</tr>
<tr>
<td>Equivalent Relations to ILI (WN1.5)</td>
<td>968</td>
<td>894</td>
<td></td>
<td>1862</td>
</tr>
<tr>
<td>Average per synset</td>
<td>1.30</td>
<td>1.92</td>
<td>0.01</td>
<td>1.46</td>
</tr>
</tbody>
</table>
Table 15: Language Internal Relations GB

<table>
<thead>
<tr>
<th>Language Internal Relations</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Others</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synsets</td>
<td>968</td>
<td>894</td>
<td>1862</td>
<td></td>
</tr>
<tr>
<td>CAUSES</td>
<td>24</td>
<td>242</td>
<td>266</td>
<td></td>
</tr>
<tr>
<td>HAS_XPOS_HYPERONYM</td>
<td>10</td>
<td>12</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>HAS_XPOS_HYPONYM</td>
<td>12</td>
<td>10</td>
<td>22</td>
<td></td>
</tr>
<tr>
<td>HAS_SUBEVENT</td>
<td>6</td>
<td>4</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>XPOS_NEAR_SYNONYM</td>
<td>492</td>
<td>492</td>
<td>984</td>
<td></td>
</tr>
<tr>
<td>XPOS_NEAR_ANTONYM</td>
<td>4</td>
<td>4</td>
<td>8</td>
<td></td>
</tr>
<tr>
<td>XPOS_FUZZYNYM</td>
<td>169</td>
<td></td>
<td>169</td>
<td></td>
</tr>
<tr>
<td>HAS_DERIVED</td>
<td></td>
<td></td>
<td>1399</td>
<td>1399</td>
</tr>
<tr>
<td>INVOLVED_AGENT</td>
<td>274</td>
<td></td>
<td>274</td>
<td></td>
</tr>
<tr>
<td>INVOLVED_TARGET_DIRECTION</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>INVOLVED_INSTRUMENT</td>
<td>104</td>
<td></td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>INVOLVED_PATIENT</td>
<td>31</td>
<td></td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>INVOLVED</td>
<td>13</td>
<td></td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>IS_CAUSED_BY</td>
<td>242</td>
<td>24</td>
<td>266</td>
<td></td>
</tr>
<tr>
<td>IS_SUBEVENT_OF</td>
<td>4</td>
<td>6</td>
<td>10</td>
<td></td>
</tr>
<tr>
<td>ROLE</td>
<td>13</td>
<td></td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>ROLE_AGENT</td>
<td>274</td>
<td></td>
<td>274</td>
<td></td>
</tr>
<tr>
<td>ROLE_INSTRUMENT</td>
<td>104</td>
<td></td>
<td>104</td>
<td></td>
</tr>
<tr>
<td>ROLE_PATIENT</td>
<td>31</td>
<td></td>
<td>31</td>
<td></td>
</tr>
<tr>
<td>ROLE_TARGET_DIRECTION</td>
<td>1</td>
<td></td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>DERIVED_FROM</td>
<td>1399</td>
<td></td>
<td>1399</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>2785</td>
<td>2616</td>
<td>5401</td>
<td></td>
</tr>
<tr>
<td>Average per synset</td>
<td>2.87</td>
<td>2.92</td>
<td>2.89</td>
<td></td>
</tr>
</tbody>
</table>

Table 16: Equivalence Relations GB

<table>
<thead>
<tr>
<th>Equivalence Relations</th>
<th>Nouns</th>
<th>Verbs</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>EQ_SYNONYM</td>
<td>968</td>
<td>894</td>
<td>1862</td>
</tr>
<tr>
<td>Total</td>
<td>968</td>
<td>894</td>
<td>1862</td>
</tr>
</tbody>
</table>

2.5 Quantitative conclusions

The total size of the wordnets aimed at is 25,000 synsets, about 50,000 word senses (synset variants) and 20,000 entries. The next table shows that the figure for the number of synsets has already been reached for Italian and Spanish: 24,207 and 21,179 synsets respectively. The coverage of the Dutch wordnet is much lower (about 50%) but still within the limit which was set for the first subset: 10,000 synsets. This difference only applies to nouns, the verbs are covered equally well in all 3 sites. The main reason for the lower coverage of nouns in the Dutch wordnet is the fact that only those synsets are included that have been processed manually, encoding a maximum of relations. In fact, a much larger Dutch fragment can be provided with hyponymy, synonymy and equivalence relations but this information needs to be verified first. In general, we can thus conclude that the project is advancing the original planning for the first subset. The remaining work will therefore not focus on extending the size of the wordnets but on improving the quality and the overlap across the wordnets (see below).

With respect to the quality, we can already draw some conclusion from table 17. First of all we see that the distribution of senses per synset and per entry is very different for each site. The Spanish synsets contain more variants (double compared to Italian) and also more senses per entry. Since they expanded the WordNet1.5 synsets with Spanish translations, these figures reflect the WordNet1.5 distribution. WordNet1.5 uses a wider notion of synonymy and a more fine-grained differentiation of senses than the traditional dictionaries on which the Italian and Dutch wordnets are based.
Table 17: First Subset Overview: NL, ES, IT

<table>
<thead>
<tr>
<th>Synsets</th>
<th>Noun</th>
<th>Verb</th>
<th>Oth.</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Synsets</td>
<td>5917</td>
<td>3282</td>
<td>389</td>
<td>9588</td>
</tr>
<tr>
<td>Number of senses</td>
<td>10874</td>
<td>5915</td>
<td>1198</td>
<td>19787</td>
</tr>
<tr>
<td>Senses per synset</td>
<td>1.84</td>
<td>1.80</td>
<td>3.08</td>
<td>1.88</td>
</tr>
<tr>
<td>Entries</td>
<td>9555</td>
<td>4211</td>
<td>1070</td>
<td>14836</td>
</tr>
<tr>
<td>Senses / entry</td>
<td>1.14</td>
<td>1.40</td>
<td>1.12</td>
<td>1.21</td>
</tr>
<tr>
<td>Language Internal Rels.</td>
<td>16917</td>
<td>9486</td>
<td>432</td>
<td>26835</td>
</tr>
<tr>
<td>LI Rels/synset</td>
<td>2.86</td>
<td>2.89</td>
<td>1.11</td>
<td>2.80</td>
</tr>
<tr>
<td>Equivalent Rels to ILI</td>
<td>7664</td>
<td>6296</td>
<td>5</td>
<td>13965</td>
</tr>
<tr>
<td>Eq. Rels/synset</td>
<td>1.30</td>
<td>1.92</td>
<td>0.01</td>
<td>1.46</td>
</tr>
<tr>
<td>Synsets without ILI</td>
<td>1578</td>
<td>394</td>
<td>385</td>
<td>2357</td>
</tr>
</tbody>
</table>

The differences in language-internal and equivalence relations per synset indicate a further difference in quality. The Dutch wordnet has the highest average of language-internal relations and the Spanish wordnet has the most equivalence relations. In fact, the Spanish equivalence-matching is 1:1 because of the followed procedure. Because they include synsets that can be translated from WordNet1.5, there are no synsets without ILI-references. The next overview tables show more details on these differences.

Table 18: Overview of Language Internal Relations

<table>
<thead>
<tr>
<th>Language Internal Relations</th>
<th>Dutch</th>
<th>Italian</th>
<th>Spanish</th>
<th>English</th>
</tr>
</thead>
<tbody>
<tr>
<td>HAS_HYPERONYM</td>
<td>9757</td>
<td>20860</td>
<td>20737</td>
<td>97.91%</td>
</tr>
<tr>
<td>HAS_HYPONYM</td>
<td>9757</td>
<td>20860</td>
<td>20737</td>
<td>97.91%</td>
</tr>
<tr>
<td>HAS_XPOS_HYPERONYM</td>
<td>48</td>
<td>2</td>
<td>0.009%</td>
<td>1.18%</td>
</tr>
<tr>
<td>HAS_XPOS_HYPONYM</td>
<td>48</td>
<td>2</td>
<td>0.009%</td>
<td>1.18%</td>
</tr>
<tr>
<td>HAS_HOLONYM</td>
<td>1074</td>
<td>331</td>
<td>1.40%</td>
<td>1366</td>
</tr>
<tr>
<td>HAS_MERONYM</td>
<td>1084</td>
<td>839</td>
<td>3.56%</td>
<td>1366</td>
</tr>
<tr>
<td>INVOLVED</td>
<td>1189</td>
<td>509</td>
<td>2.16%</td>
<td>6</td>
</tr>
<tr>
<td>ROLE</td>
<td>836</td>
<td>147</td>
<td>0.62%</td>
<td>6</td>
</tr>
<tr>
<td>CAUSES</td>
<td>749</td>
<td>468</td>
<td>1.99%</td>
<td>40</td>
</tr>
<tr>
<td>IS_CAUSED_BY</td>
<td>609</td>
<td>425</td>
<td>1.80%</td>
<td>40</td>
</tr>
<tr>
<td>HAS_SUBEVENT</td>
<td>208</td>
<td>34</td>
<td>0.14%</td>
<td>1</td>
</tr>
<tr>
<td>IS_SUBEVENT_OF</td>
<td>219</td>
<td>1</td>
<td>0.004%</td>
<td>1</td>
</tr>
<tr>
<td>NEAR_ANTONYM</td>
<td>349</td>
<td>20</td>
<td>0.08%</td>
<td></td>
</tr>
<tr>
<td>NEAR_SYNONYM</td>
<td>219</td>
<td>225</td>
<td>0.95%</td>
<td>6</td>
</tr>
<tr>
<td>BE_IN_STATE</td>
<td>93</td>
<td>123</td>
<td>0.52%</td>
<td></td>
</tr>
<tr>
<td>STATE_OF</td>
<td>94</td>
<td>99%</td>
<td>0.98%</td>
<td></td>
</tr>
<tr>
<td>XPOS_NEAR_ANTONYM</td>
<td>6</td>
<td>984</td>
<td>52.84%</td>
<td></td>
</tr>
<tr>
<td>XPOS_NEAR_SYNONYM</td>
<td>496</td>
<td>9082</td>
<td>38.60%</td>
<td>2</td>
</tr>
<tr>
<td>HAS_DERIVED</td>
<td>1399</td>
<td>44308</td>
<td>75.13%</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>26835</td>
<td>57091</td>
<td>44308</td>
<td>5401</td>
</tr>
<tr>
<td>Synsets</td>
<td>9588</td>
<td>23523</td>
<td>21179</td>
<td>1862</td>
</tr>
<tr>
<td>Rels/Synset</td>
<td>2.80</td>
<td>2.42</td>
<td>2.09</td>
<td>2.89</td>
</tr>
</tbody>
</table>

The first column gives the absolute number of relations per type, the second column for each language gives the relative percentage of the relation for all the covered synsets. Except for English, hyponymy is almost 100% covered. This means that each synset has at least one hyperonym average. For the rest, we see that the Dutch wordnet incorporates far more other relations than the other wordnets. This is in line with the strategy followed for the Dutch.

---

1 In this table we did not include the figure from SHE because their subset is too different to be compared.
2 The hyperonym relation is more than 100% because synsets may have multiple hyperonyms.
wordnet, to focus on the rich encoding of the most important concepts rather than a large coverage with shallow information. In the case of \textbf{XPOS\_NEAR\_SYNONYM} we see an extreme number of relations for Italian. The English coverage of relations is very different, since they focus on adding XPOS relations missing in WordNet1.5.

\textit{Table 19: Overview of Equivalence Relations}

| Equivalence Relations | Dutch | | | Spanish | | | | Italian | | |
|-----------------------|-------|---|---|-------|---|---|-------|---|---|---|---|
|                       | Nouns | Verbs | Total | Nouns | Verbs | Total | Nouns | Verbs | Total | Nouns | Verbs | Total |
| EQ\_SYNONYM           | 1370  | 375  | 1745  | 18577 | 2602  | 21179 | 3697  | 307  | 4004 |
| EQ\_NEAR\_SYNONYM     | 6025  | 5883 | 11908 |       |       |       | 631   | 259  | 890  |
| EQ\_HAS\_HYPERONYM    | 174   | 22   | 196   | 40    |       | 1947  | 77    | 2024 |
| EQ\_HAS\_HYPONYM      | 85    | 11   | 96    | 14    |       | 10    | 7     | 18   |
| EQ\_INVOLVED          |       | 4    | 4     |       |       |       | 1     | 1    | 2    |
| EQ\_IS\_ CAUSED\_BY   | 1     | 1    | 2     |       |       |       |       |       |      |
| EQ\_HAS\_HOLONYM      | 3     | 3    | 1     | 1     |       |       |       |       |      |
| EQ\_HAS\_MERONYM      | 6     | 6    | 2     | 2     |       |       |       |       |      |
| Total                 | 7664  | 6296 | 13960 | 18634 | 2602  | 21236 | 6275  | 653  | 6928 |

Equivalence relations for most of the Dutch nouns and verbs and most of the Spanish nouns are generated automatically. In the Dutch wordnet, the automatically generated equivalences are always of the type EQ\_NEAR\_SYNONYM, which explains the high figure. All other equivalences are encoded manually. The equivalences for the Spanish verb are all created manually. Equivalence relations for the Italian synsets are generated semi-automatically but are all manually verified. The best-3 equivalences have been chosen for the Dutch wordnet, which explains the high average of equivalence relations. The main work to be done especially for Dutch and also for Spanish is to improve the quality of the equivalence relations. For the Italian wordnets, the quantity of equivalences has to be increased.

\textit{2.6 Overlap with Parole lexicons}

The aim of the Parole project is to develop morpho-syntactic lexicons for the most frequent words of the European languages. As such Parole is complementary to EuroWordNet. Future developers should be able to take the generic resources of EuroWordNet and Parole to develop combined lexicons for their NLP applications. It is therefore important to make sure that more or less the same vocabulary is covered in both projects. In both projects the most frequent words should be represented. We therefore compared the lexicons in EuroWordNet with Parole for different corpus frequencies. However, for Dutch and Italian that Parole data are not yet available. For Dutch we therefore used the Celex frequency information, which is based on a 40MLN token corpus.

\textit{Table 20: Coverage of Dutch Subset1 related to INL/Celex frequency}

| Frequency | Celex entries | Nouns | | Verbs | | |
|-----------|---------------|-------|---|---|---|---|---|---|---|
|           |               | Celex covered | %coverage | Celex entries | | Celex covered | %coverage |
| 1001-     | 1217          | 910   | 74.77% | 677 | 597 | 88.18% |
| 501-1000  | 939           | 449   | 47.82% | 455 | 315 | 69.23% |
| 251-500   | 1408          | 509   | 36.15% | 637 | 391 | 61.38% |
| 101-250   | 3157          | 893   | 28.29% | 1176| 642 | 54.59% |
| 51-100    | 3604          | 748   | 20.75% | 957 | 440 | 45.98% |
| 31-50     | 3380          | 565   | 16.72% | 695 | 271 | 38.99% |
| 21-30     | 3016          | 477   | 15.82% | 495 | 191 | 38.59% |
| 11-20     | 5258          | 722   | 13.73% | 706 | 265 | 37.54% |
| 6-10      | 4804          | 550   | 11.45% | 567 | 212 | 37.39% |
| 3-5       | 4713          | 505   | 10.72% | 377 | 135 | 35.81% |
| 2         | 2338          | 229   | 9.79%  | 346 | 113 | 32.66% |
| 0         | 127           | 25    | 19.69% | 9   | 2   | 22.22% |
| 1         | 30001         | 2885  | 9.62%  | 1725| 499 | 28.93% |
| overall   | 63962         | 9467  | 14.80% | 8822| 4073| 46.17% |
Table 21: Coverage of Spanish Subset1 of Parole Lexicons

<table>
<thead>
<tr>
<th>Frequency</th>
<th>parole entries</th>
<th>parole covered</th>
<th>% coverage</th>
<th>parole entries</th>
<th>parole covered</th>
<th>% coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001-</td>
<td>147</td>
<td>143</td>
<td>97.28</td>
<td>110</td>
<td>107</td>
<td>97.27</td>
</tr>
<tr>
<td>501-1000</td>
<td>261</td>
<td>246</td>
<td>94.25</td>
<td>139</td>
<td>118</td>
<td>84.89</td>
</tr>
<tr>
<td>251-500</td>
<td>462</td>
<td>429</td>
<td>92.86</td>
<td>218</td>
<td>172</td>
<td>78.90</td>
</tr>
<tr>
<td>101-250</td>
<td>933</td>
<td>863</td>
<td>92.50</td>
<td>381</td>
<td>257</td>
<td>67.45</td>
</tr>
<tr>
<td>51-100</td>
<td>959</td>
<td>863</td>
<td>89.99</td>
<td>374</td>
<td>265</td>
<td>70.86</td>
</tr>
<tr>
<td>31-50</td>
<td>892</td>
<td>804</td>
<td>90.13</td>
<td>347</td>
<td>185</td>
<td>53.31</td>
</tr>
<tr>
<td>21-30</td>
<td>730</td>
<td>632</td>
<td>86.57</td>
<td>286</td>
<td>141</td>
<td>49.30</td>
</tr>
<tr>
<td>11-20</td>
<td>1202</td>
<td>978</td>
<td>81.36</td>
<td>469</td>
<td>175</td>
<td>37.31</td>
</tr>
<tr>
<td>6-10</td>
<td>1024</td>
<td>790</td>
<td>77.15</td>
<td>360</td>
<td>129</td>
<td>35.83</td>
</tr>
<tr>
<td>3-5</td>
<td>968</td>
<td>665</td>
<td>68.70</td>
<td>254</td>
<td>74</td>
<td>29.13</td>
</tr>
<tr>
<td>2</td>
<td>435</td>
<td>257</td>
<td>59.08</td>
<td>123</td>
<td>32</td>
<td>26.02</td>
</tr>
<tr>
<td>1</td>
<td>643</td>
<td>334</td>
<td>51.94</td>
<td>131</td>
<td>26</td>
<td>19.85</td>
</tr>
<tr>
<td>overall</td>
<td>8656</td>
<td>7004</td>
<td>80.91</td>
<td>3192</td>
<td>1681</td>
<td>52.66</td>
</tr>
</tbody>
</table>

These tables show that there is a very high overlap for the higher frequencies. This is according to our expectation that frequent words are also relatively general and basic and therefore are likely to be occur among the Base Concepts. Each site will individually extract the missing top-frequent words and integrate them in the next building phase.

Finally, the next table shows how the WordNet1.5 matches with the English Parole lexicon, where the frequency information is based on the Cobuild frequency information in CELEX. The coverage for most frequencies is very high. There is a strange deviation for verbs with frequencies 3-5 for which we do not have an explanation.

Table 22: Coverage of WordNet1.5 compared to the English Parole Lexicon

<table>
<thead>
<tr>
<th>Frequency</th>
<th>parole entries</th>
<th>parole covered</th>
<th>% coverage</th>
<th>parole entries</th>
<th>parole covered</th>
<th>% coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1001-</td>
<td>767</td>
<td>698</td>
<td>91</td>
<td>335</td>
<td>333</td>
<td>99.4</td>
</tr>
<tr>
<td>501-1000</td>
<td>675</td>
<td>604</td>
<td>89.5</td>
<td>251</td>
<td>250</td>
<td>99.6</td>
</tr>
<tr>
<td>251-500</td>
<td>947</td>
<td>863</td>
<td>91.1</td>
<td>366</td>
<td>366</td>
<td>100</td>
</tr>
<tr>
<td>101-250</td>
<td>1677</td>
<td>1680</td>
<td>99.8</td>
<td>681</td>
<td>681</td>
<td>100</td>
</tr>
<tr>
<td>51-100</td>
<td>1556</td>
<td>1571</td>
<td>99</td>
<td>729</td>
<td>729</td>
<td>100</td>
</tr>
<tr>
<td>31-50</td>
<td>1376</td>
<td>1376</td>
<td>100</td>
<td>516</td>
<td>516</td>
<td>100</td>
</tr>
<tr>
<td>21-30</td>
<td>1090</td>
<td>1091</td>
<td>99.9</td>
<td>332</td>
<td>332</td>
<td>100</td>
</tr>
<tr>
<td>11-20</td>
<td>1525</td>
<td>1529</td>
<td>99.8</td>
<td>392</td>
<td>392</td>
<td>100</td>
</tr>
<tr>
<td>6-10</td>
<td>1024</td>
<td>1024</td>
<td>100</td>
<td>160</td>
<td>111</td>
<td>69.3</td>
</tr>
<tr>
<td>3-5</td>
<td>650</td>
<td>650</td>
<td>100</td>
<td>52</td>
<td>17</td>
<td>32.7</td>
</tr>
<tr>
<td>2</td>
<td>218</td>
<td>218</td>
<td>100</td>
<td>15</td>
<td>15</td>
<td>100</td>
</tr>
<tr>
<td>1</td>
<td>157</td>
<td>157</td>
<td>100</td>
<td>18</td>
<td>18</td>
<td>100</td>
</tr>
<tr>
<td>0 in CELEX</td>
<td>171</td>
<td>171</td>
<td>100</td>
<td>38</td>
<td>38</td>
<td>100</td>
</tr>
<tr>
<td>not in CELEX</td>
<td>676</td>
<td>558</td>
<td>82.5</td>
<td>356</td>
<td>287</td>
<td>80.6</td>
</tr>
<tr>
<td>overall</td>
<td>12509</td>
<td>12190</td>
<td>97.4</td>
<td>4241</td>
<td>4085</td>
<td>96.3</td>
</tr>
</tbody>
</table>
2.7 Coverage of Subset1 over top concept clusters

As explained in the introduction, 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. To still get an idea of the conceptual distribution of this extension we also measure the progress of the wordnets relative to the Top Ontology, which represents the diversity of Base Concepts that have been selected. For this purpose, AMS implemented an inheritance mechanism that derives the Top Concepts from hyperonyms in WordNet1.5. By loading ILLI-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 in the Italian and Dutch wordnets have a (correct) equivalent relation. Furthermore, it may be that the hyponymy relations in the local wordnets are different, but the global semantic classification still has to be consistent. This method therefore still gives a good indication of the conceptual coverage.

The Top Ontology is divided in 3 main parts:

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

However, there are cases where the hyponymy links do not provide any top-concept: i.e. not all WordNet1.5 tops have been classified.

Table 23: Synsets that are not clustered by the Top Ontology

<table>
<thead>
<tr>
<th>VOID</th>
<th>WN</th>
<th>NL</th>
<th>ES</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>nouns</td>
<td>0</td>
<td>17</td>
<td>33</td>
<td>15</td>
</tr>
<tr>
<td>verbs</td>
<td>2109</td>
<td>310</td>
<td>638</td>
<td>385</td>
</tr>
</tbody>
</table>

WordNet1.5 only has 11 tops for Nouns but 573 for verbs. Most of the noun tops have at least one Top Concept assigned, whereas only 48 of the verb tops have been classified so far. This explains that only a few nominal synsets have not inherited an top concept, whereas a large proportion of the verbs is not (in)directly linked to the ontology. In the near future we will classify all the WordNet1.5 tops so that a complete clustering can be made.

Table 24: Nominal Synsets clustered as 1stOrder Concepts

<table>
<thead>
<tr>
<th>Nouns</th>
<th>WN</th>
<th>NL</th>
<th>ES</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1stOrderEntity</td>
<td>5</td>
<td>0.00</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>Animal</td>
<td>4024</td>
<td>2.76</td>
<td>0.59</td>
<td>0.04</td>
</tr>
<tr>
<td>Artifact</td>
<td>12054</td>
<td>8.27</td>
<td>12.86</td>
<td>0.82</td>
</tr>
<tr>
<td>Building</td>
<td>589</td>
<td>0.40</td>
<td>1.13</td>
<td>0.07</td>
</tr>
<tr>
<td>Comestible</td>
<td>2207</td>
<td>151</td>
<td>1.65</td>
<td>0.11</td>
</tr>
<tr>
<td>Container</td>
<td>1060</td>
<td>0.73</td>
<td>0.63</td>
<td>0.04</td>
</tr>
<tr>
<td>Covering</td>
<td>1279</td>
<td>0.88</td>
<td>1.11</td>
<td>0.07</td>
</tr>
<tr>
<td>Creature</td>
<td>473</td>
<td>0.32</td>
<td>0.02</td>
<td>0.00</td>
</tr>
<tr>
<td>Function</td>
<td>10183</td>
<td>6.99</td>
<td>5.78</td>
<td>0.37</td>
</tr>
<tr>
<td>Functional</td>
<td>120</td>
<td>0.08</td>
<td>0.19</td>
<td>0.01</td>
</tr>
<tr>
<td>Furniture</td>
<td>196</td>
<td>0.13</td>
<td>0.18</td>
<td>0.01</td>
</tr>
<tr>
<td>Garment</td>
<td>446</td>
<td>0.31</td>
<td>0.24</td>
<td>0.02</td>
</tr>
<tr>
<td>Gas</td>
<td>56</td>
<td>0.04</td>
<td>0.09</td>
<td>0.01</td>
</tr>
<tr>
<td>Group</td>
<td>13092</td>
<td>8.98</td>
<td>225</td>
<td>2.42</td>
</tr>
<tr>
<td>Human</td>
<td>6315</td>
<td>4.33</td>
<td>215</td>
<td>2.31</td>
</tr>
<tr>
<td>ImageRepresentation</td>
<td>480</td>
<td>0.33</td>
<td>28</td>
<td>0.30</td>
</tr>
<tr>
<td>Instrument</td>
<td>4557</td>
<td>3.13</td>
<td>512</td>
<td>5.50</td>
</tr>
<tr>
<td>Language Representation</td>
<td>1883</td>
<td>1.29</td>
<td>107</td>
<td>1.15</td>
</tr>
<tr>
<td>-------------------------</td>
<td>------</td>
<td>------</td>
<td>-----</td>
<td>------</td>
</tr>
<tr>
<td>Liquid</td>
<td>1083</td>
<td>0.74</td>
<td>67</td>
<td>0.72</td>
</tr>
<tr>
<td>Living</td>
<td>16375</td>
<td>11.23</td>
<td>484</td>
<td>5.20</td>
</tr>
<tr>
<td>Money Representation</td>
<td>241</td>
<td>0.17</td>
<td>23</td>
<td>0.25</td>
</tr>
<tr>
<td>Natural</td>
<td>15182</td>
<td>10.41</td>
<td>1646</td>
<td>17.67</td>
</tr>
<tr>
<td>Object</td>
<td>26174</td>
<td>17.96</td>
<td>1864</td>
<td>20.01</td>
</tr>
<tr>
<td>Occupation</td>
<td>1222</td>
<td>0.84</td>
<td>42</td>
<td>0.45</td>
</tr>
<tr>
<td>Part</td>
<td>7412</td>
<td>5.08</td>
<td>587</td>
<td>6.30</td>
</tr>
<tr>
<td>Place</td>
<td>3235</td>
<td>2.22</td>
<td>220</td>
<td>2.36</td>
</tr>
<tr>
<td>Plant</td>
<td>5619</td>
<td>3.85</td>
<td>81</td>
<td>0.87</td>
</tr>
<tr>
<td>Representation</td>
<td>592</td>
<td>0.41</td>
<td>55</td>
<td>0.59</td>
</tr>
<tr>
<td>Software</td>
<td>134</td>
<td>0.09</td>
<td>5</td>
<td>0.05</td>
</tr>
<tr>
<td>Solid</td>
<td>3985</td>
<td>2.73</td>
<td>324</td>
<td>3.48</td>
</tr>
<tr>
<td>Substance</td>
<td>5045</td>
<td>3.46</td>
<td>514</td>
<td>5.52</td>
</tr>
<tr>
<td>Vehicle</td>
<td>453</td>
<td>0.31</td>
<td>38</td>
<td>0.41</td>
</tr>
<tr>
<td>Total</td>
<td>145771</td>
<td>9316</td>
<td>639</td>
<td>43358</td>
</tr>
</tbody>
</table>

The first column gives the full list of the 1stOrder Top Concepts. The first column of each wordnet gives the number of synsets (represented as ILI-records) that are either directly or indirectly via a hyperonym chain classified by the Top-Concept. The next column gives the percentage of the total set of 1stOrder nouns covered by each wordnet and the third column for NL, ES and IT gives the percentage of the total set in WordNet1.5. The second columns of each wordnet gives the distribution per top-concept class. 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 should be extended in this domain to become more balanced.5

---

5 The table is also useful for users of the wordnets to verify if particular domains or fields of their interest are well-represented or need to be extended.
In this table we clearly see that the Spanish wordnet closely follows the balancing of WordNet1.5, due to the methodology that has been applied. The Dutch and Italian wordnets show a diverging distribution. The fields with relatively lower coverage are marked in the table:

Spanish wordnet: Group.
Dutch wordnet: Animal; Creature; Group; Human; Living; Occupation; Plant.
Italian wordnet: Building; Container; Covering; Furniture; Garment; LanguageRepresentation; Part; Place; Solid; Substance.

The differences in distribution do not necessarily imply that the areas are badly represented. They can also be due to differences in lexicalization across the languages or to a lack of equivalence relations in a particular area. Nevertheless, each wordnet builder has to check these fields in their resources to find whether these differences are due to incompleteness or due to lexicalization differences. In the former case, the wordnets have to be extended. Note that regardless of the balancing or distribution of the synsets, the total coverage is much lower than WordNet1.5 and should especially be increased for the Dutch wordnet.

The next two tables shows the distribution for nouns and verbs that are classified as 2ndOrderEntities according to the WordNet1.5 hyponymy chains. Whereas the previous table showed some differences in conceptual coverage, the next tables are remarkably balanced. Only Quantity and Usage are slightly under-represented in the Dutch wordnet and latter in the Italian wordnet.

Table 25: Nominal Synsets clustered as 2ndOrder Concepts

<table>
<thead>
<tr>
<th>Nouns</th>
<th>WNL5</th>
<th>NL</th>
<th>ES</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agentive</td>
<td>6146</td>
<td>6,96</td>
<td>324</td>
<td>5,72</td>
</tr>
<tr>
<td>BoundedEvent</td>
<td>4753</td>
<td>5,38</td>
<td>292</td>
<td>5,16</td>
</tr>
<tr>
<td>Cause</td>
<td>5496</td>
<td>6,23</td>
<td>314</td>
<td>5,55</td>
</tr>
<tr>
<td>Communication</td>
<td>3852</td>
<td>4,36</td>
<td>223</td>
<td>3,94</td>
</tr>
<tr>
<td>Condition</td>
<td>2325</td>
<td>2,63</td>
<td>311</td>
<td>5,49</td>
</tr>
<tr>
<td>Dynamic</td>
<td>9400</td>
<td>10,65</td>
<td>610</td>
<td>10,78</td>
</tr>
<tr>
<td>Existence</td>
<td>198</td>
<td>0,22</td>
<td>19</td>
<td>0,34</td>
</tr>
<tr>
<td>Experience</td>
<td>4012</td>
<td>4,55</td>
<td>294</td>
<td>5,19</td>
</tr>
<tr>
<td>Location</td>
<td>851</td>
<td>0,96</td>
<td>60</td>
<td>1,06</td>
</tr>
<tr>
<td>Manner</td>
<td>573</td>
<td>0,65</td>
<td>29</td>
<td>0,51</td>
</tr>
<tr>
<td>Mental</td>
<td>6166</td>
<td>6,99</td>
<td>390</td>
<td>6,89</td>
</tr>
<tr>
<td>Modal</td>
<td>291</td>
<td>0,33</td>
<td>17</td>
<td>0,30</td>
</tr>
<tr>
<td>Phenomenal</td>
<td>1216</td>
<td>1,38</td>
<td>144</td>
<td>2,54</td>
</tr>
<tr>
<td>Physical</td>
<td>4712</td>
<td>5,34</td>
<td>445</td>
<td>7,86</td>
</tr>
<tr>
<td>Possession</td>
<td>95</td>
<td>0,11</td>
<td>6</td>
<td>0,11</td>
</tr>
<tr>
<td>Property</td>
<td>6975</td>
<td>7,90</td>
<td>505</td>
<td>8,92</td>
</tr>
<tr>
<td>Purpose</td>
<td>9250</td>
<td>10,48</td>
<td>459</td>
<td>8,11</td>
</tr>
<tr>
<td>Quantity</td>
<td>2129</td>
<td>2,41</td>
<td>95</td>
<td>1,68</td>
</tr>
<tr>
<td>Relation</td>
<td>4148</td>
<td>4,70</td>
<td>248</td>
<td>4,38</td>
</tr>
<tr>
<td>Social</td>
<td>7449</td>
<td>8,44</td>
<td>353</td>
<td>6,24</td>
</tr>
<tr>
<td>Static</td>
<td>3408</td>
<td>3,86</td>
<td>310</td>
<td>5,48</td>
</tr>
<tr>
<td>Stimulating</td>
<td>599</td>
<td>0,68</td>
<td>42</td>
<td>0,74</td>
</tr>
<tr>
<td>Time</td>
<td>712</td>
<td>0,81</td>
<td>32</td>
<td>0,57</td>
</tr>
<tr>
<td>UnboundedEvent</td>
<td>2792</td>
<td>3,16</td>
<td>124</td>
<td>2,19</td>
</tr>
<tr>
<td>Usage</td>
<td>723</td>
<td>0,82</td>
<td>14</td>
<td>0,25</td>
</tr>
<tr>
<td>Total</td>
<td>88271</td>
<td>5660</td>
<td>6,41</td>
<td>35085</td>
</tr>
</tbody>
</table>
**Table 26: Verbal Synsets clustered as 2ndOrder Concepts**

<table>
<thead>
<tr>
<th>Verbs</th>
<th>UN</th>
<th>NL</th>
<th>ES</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agentive</td>
<td>3139</td>
<td>9.15</td>
<td>344</td>
<td>7.86</td>
</tr>
<tr>
<td>BoundedEvent</td>
<td>4038</td>
<td>11.77</td>
<td>496</td>
<td>11.33</td>
</tr>
<tr>
<td>Cause</td>
<td>3265</td>
<td>9.52</td>
<td>422</td>
<td>9.64</td>
</tr>
<tr>
<td>Communication</td>
<td>1067</td>
<td>3.11</td>
<td>123</td>
<td>2.81</td>
</tr>
<tr>
<td>Condition</td>
<td>578</td>
<td>1.69</td>
<td>65</td>
<td>1.48</td>
</tr>
<tr>
<td>Dynamic</td>
<td>5856</td>
<td>17.07</td>
<td>892</td>
<td>20.37</td>
</tr>
<tr>
<td>Existence</td>
<td>831</td>
<td>2.42</td>
<td>89</td>
<td>2.03</td>
</tr>
<tr>
<td>Experience</td>
<td>384</td>
<td>1.12</td>
<td>46</td>
<td>1.05</td>
</tr>
<tr>
<td>Location</td>
<td>2579</td>
<td>7.52</td>
<td>408</td>
<td>9.32</td>
</tr>
<tr>
<td>Manner</td>
<td>174</td>
<td>0.51</td>
<td>21</td>
<td>0.48</td>
</tr>
<tr>
<td>Mental</td>
<td>840</td>
<td>2.45</td>
<td>75</td>
<td>1.71</td>
</tr>
<tr>
<td>Modal</td>
<td>16</td>
<td>0.05</td>
<td>2</td>
<td>0.05</td>
</tr>
<tr>
<td>Phenomenal</td>
<td>10</td>
<td>0.03</td>
<td>3</td>
<td>0.07</td>
</tr>
<tr>
<td>Physical</td>
<td>2938</td>
<td>8.57</td>
<td>328</td>
<td>7.49</td>
</tr>
<tr>
<td>Possession</td>
<td>655</td>
<td>1.91</td>
<td>88</td>
<td>2.01</td>
</tr>
<tr>
<td>Property</td>
<td>170</td>
<td>0.50</td>
<td>14</td>
<td>0.32</td>
</tr>
<tr>
<td>Purpose</td>
<td>1896</td>
<td>5.53</td>
<td>189</td>
<td>4.32</td>
</tr>
<tr>
<td>Quantity</td>
<td>302</td>
<td>0.88</td>
<td>40</td>
<td>0.91</td>
</tr>
<tr>
<td>Relation</td>
<td>307</td>
<td>0.90</td>
<td>36</td>
<td>0.82</td>
</tr>
<tr>
<td>SituationType</td>
<td>73</td>
<td>0.21</td>
<td>14</td>
<td>0.32</td>
</tr>
<tr>
<td>Social</td>
<td>1391</td>
<td>4.06</td>
<td>157</td>
<td>3.59</td>
</tr>
<tr>
<td>Static</td>
<td>165</td>
<td>0.48</td>
<td>24</td>
<td>0.55</td>
</tr>
<tr>
<td>Stimulating</td>
<td>334</td>
<td>0.97</td>
<td>24</td>
<td>0.55</td>
</tr>
<tr>
<td>Time</td>
<td>14</td>
<td>0.04</td>
<td>0</td>
<td>0.00</td>
</tr>
<tr>
<td>UnboundedEvent</td>
<td>1037</td>
<td>3.02</td>
<td>145</td>
<td>3.31</td>
</tr>
<tr>
<td>Usage</td>
<td>129</td>
<td>0.38</td>
<td>24</td>
<td>0.55</td>
</tr>
<tr>
<td>Total</td>
<td>34297</td>
<td>97.17</td>
<td>4379</td>
<td>12.77</td>
</tr>
</tbody>
</table>

The fact that the 2ndOrderEntities are equally balanced may also indicate that the Top-Ontology classification is more shallow compared to the nominal classification. A shallow, more abstract classification necessarily tends to blur out differences as well. From these tables, we nevertheless cannot derive any conclusions for extending the wordnets in a particular direction.

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 that the coverage for Spanish and Dutch is similar to the total coverage of 1stOrder and 2ndOrder Entities compared to WordNet1.5. The Italian wordnet however shows a significantly lower percentage here. This can again either be due to lexicalization differences, incompleteness or to a lack of equivalence relations in this area.

**Table 27: Nominal Synsets clustered as 3rdOrder Concepts**

<table>
<thead>
<tr>
<th>3rdOrderEntity</th>
<th>UN</th>
<th>NL</th>
<th>ES</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>D014D015: The Restructured Core wordnets in EuroWordNet: Subset1</td>
<td>4989</td>
<td>309</td>
<td>6.19%</td>
<td>1860</td>
</tr>
</tbody>
</table>
3. Comparison of the first Subset

We have carried out two different types of comparisons for Subset1:

- an in-depth comparison of the wordnets in the EuroWordNet database for 18 semantic fields
- an overall comparison of the full subset

The in-depth-comparison is carried out using the comparison options in the Polaris tool. The overall comparison is done by generating the hyperonym chains for the full subset in the form of the ILI-records. This resulted in compatible graph-structures for each wordnet. FUE has developed a special toolkit for comparing these graph-structures.

3.1 Comparing specific semantic fields in the EuroWordNet database

The goal of the comparison is to measure the quality and quantity of the local wordnet by comparing them to the other wordnets (see [Peters et al., fc] for further details). For the comparison, we make a distinction between the Source wordnet and the Reference wordnets. The Source wordnet is the wordnet at a local site which is going to be evaluated by comparing it to the Reference wordnets. It is not the purpose to evaluate the Reference wordnets. A comparison will give information on:

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

The following Clusters have been examined for the Subset1 of nouns and verbs, where a division is made between 1stOrderEntities (FOEs) and 2nd/3rdOrderEntities (HighOrderEntities or HOEs):

AMS FOE Building, Comestible, Container, Covering
    HOE Feelings, Phenomena
FUE FOE Garment, Place, Furniture, Plant
    HOE Cooking, Sounds
PSA FOE Animal, Human, Instrument, Vehicle
    HOE Movements, Knowledge

Each site will distribute the major hyperonyms that represent the most important tops of these semantic fields, e.g.: {construction-4} in WordNet1.5, {bouwwerk-1} in Dutch, {construzione-1} in Italian and {construcción-4} in Spanish. The comparison then globally consists of:

- Extract the hyponyms of the Representative hyperonyms in these fields in each wordnet.
- Project the hyponyms of the Reference wordnets to the Source wordnet.
- Compare the projected hyponyms with the hyponyms in the Source wordnet

The projections in the EuroWordNet database result in sets of word meanings (WMs) in the source wordnet related to the same Inter-Lingual-Index concepts. The next screen-dump of the EuroWordNet database (Polaris) shows such a projection from the hyponyms of construcción-4 in the Spanish wordnet (the left window) to the Dutch wordnet. In the right window the Dutch WMs are shown that are related to the same ILIs given as equivalences of the Spanish hyponyms. The bottom window shows the corresponding ILI-records.
A projection may partly overlap with the set of hyponyms build up in the local Source wordnet (the hyponyms of bouwwerk-1 in Dutch). It is possible to compare sets of WMs in the database and to derive the intersection, union and difference. The intersection represents the degree of compatibility of wordnets. The WMs which are unique in the projection of the Reference wordnets or which are unique in the Source wordnet hyponyms represent the incompatibility of the wordnets. Unique sets of WMs are thus both present in the Source wordnet (i.e. the ILI-records generated by the projection have also been used in the Source-wordnet to link local synsets to the ILI) but are classified differently. The projection which is unique in the Reference wordnet is then apparently not a hyponym (at any level) of “bouwwerk 1” (construction 4) in the Dutch wordnet. On the other hand, the ILIs linked to the WMs which are unique in the Source wordnet are not part of the set of ILIs projected by the Reference wordnets (they may be present but classified differently). Projected WMs which are Unique in the Reference wordnets can be diagnosed as follows:

- **Projected by a wrong translation in the Source wordnet**: e.g. “afsluiting 2” (the event of blocking a passage or container) is automatically but wrongly translated to an object or construction with that function “barrier 1”. The event will not show up as a hyponym of the Dutch equivalent “bouwwerk” (construction 4) but it will be generated by the projection of hyponyms of “construction 4”.  
- **Wrongly classified in the Source wordnet**: e.g. “onderdak 1” (shelter 1, place to stay) is classified as “gelegenheid 1” (occasion) which is the wrong sense. It should have been classified as “gelegenheid 2” (place or building with a purpose).
- **Alternatively classified in the Source wordnet**: e.g. “centrum 3” (center 4) is defined in Dutch as “place, institution, building, area, where certain activities take place”. It is only classified in the Dutch wordnet as “institution” and not as “building”.

A more systematic overview can be provided by generating all the hyperonyms for the WMs projected by the Reference wordnet and not included in the Source hyponym set. Below is a list of the most frequent hyperonyms in Dutch for this set:

11 eigenschap 1 (property)
12 organisatie 3 (organisation)

---

6 This will also show up when we project the Dutch hyponyms back to Dutch. In that case, WMs that are linked to these ILI-records by mistake will also be projected, but they are not part of the original set (if the hyponymy relations are correct).
The Restructured Core wordnets in EuroWordNet: Subset1

14 woning 1 (living, home)
16 groep 4 (group)
21 steun 1 (support)
27 ruimte 3 (space)
42 plaats 1 (place)
53 deel 2 (part)
70 voorwerp 1 (object)
70 zaak 1 (thing)
71 entiteit 1 (thing)
71 object 1 (object)
430 iets 1 LEAF (anything)

The final “iets” (anything) is meaningless, because it is the top of all, but there are also some meaningful hyperonyms. The cluster “woning” (living), “ruimte” (space), “plaats” (place) represents places not classified as constructions but many of these may very well get an additional hyponymy link to construction. Those linked to “organisatie” (organisation) will either be solved by so-called EQ_METONYM links to the new sense-groups (after the ILI has been extended with these global senses, see section 4) or they need an extra classification or sense for the construction/building where the institutes are settled. Note that some of these can also have wrong ILI-links. Whenever there are two senses in the Dutch wordnet, one for the building and one for the institute, they should not be translated to the same sense in the ILI.

The hyperonyms “voorwerp”, “object”, “zaak” represent physical objects which are not incompatible with constructions, but which are also not very meaningful because they represent a very large and diverse group. Something similar can be said for “deel 2” (part) and “groep 4” (group), which often refer to parts or groups of constructions. It may be the case, that these can still get an additional link to construction as well. Finally, “eigenschap” (property) is a hyperonym that is totally incompatible with constructions. These must all be errors in the translation or in the classification. The above overview cannot be generated directly by Polaris. It has to be done by either exporting the WMs, which are unique in the Reference wordnets, or by loading the senses in a local tool (which has been done here for the Dutch WMs, using the AMS LDB).

The evaluation of the differences minimally consist of a manual inspection in Polaris by looking at WMs and counting the incorrect cases: i.e. WMs that cannot possibly be constructions. Inspection of the WMs that are Unique in the Source wordnets will also be done by hand, going through the list in Polaris. It may be the case that these WMs are not covered in the Reference wordnets, or that one of the above explanations applies. However, we will not evaluate the Reference wordnets, so it suffices to browse through the list and count the number of WMs that do not belong there.

Finally, in some cases the projection of a Reference wordnet may not generated output in the Source wordnet. The unmatched IILs can be projected by taking all the senses of the variants. This generates a lot of WMs in the Source wordnet which fall outside the scope of the comparison. However, after taking the intersection of the projection on a word level with the hyponyms in the Source wordnet, it is possible to filter out near-matches. We speak of near-matches when two synsets are related to the same ILL-word but to different senses. Because of the sense-differentiation in WordNet1.5 this is a likely cause of mismatch across wordnets. The next figure gives an overview of the sets of word meanings that can be generated for hyponyms of constructions, where Dutch is the Source wordnet and the other languages represent the Reference wordnets. The remaining cases of WMS that cannot be projected fall into the following classes:

- Not included in the subset of the Source wordnet
- Gaps in the language
- Gaps in the resource
In Appendix I, each site reports on the comparison of their source wordnet for the assigned clusters by comparing it to the other Reference wordnets. Here we will summarize the results.

The selected hyperonyms that have been used to derive a semantic field are given in the next table. Each hyperonym is represented by a synset, rounded by curled brackets, where we listed only a single variant. In some cases, several hyperonyms are given to represent the field. A field could not always be represented. For example, for Dutch there is no equivalent for container. After each hyperonym or set of hyperonyms we have listed the number of (sub)-hyponyms that occur in the wordnets. Finally, in the case of PLANT and ANIMAL the sets in WordNet1.5 are extremely big. We therefore limited PLANT to the first 3 hyponymy levels only and ANIMAL to the major classes such as MAMMAL, BIRD. The selections in the other wordnets are however complete. Because of the size of the field HUMAN we have split it into two sub-fields: ARTIST and WORKER.
### Table 28: Hyperonyms in the wordnets selected for the in-depth comparison

<table>
<thead>
<tr>
<th></th>
<th>WN</th>
<th>NL</th>
<th>IT</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>BUILDING</td>
<td>{construction-4}</td>
<td>1220</td>
<td>{construzione-1}</td>
<td>194</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>{ediﬁcio-1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>{manifattura-1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>{dimora-2}</td>
<td></td>
</tr>
<tr>
<td>COMESTIBLE</td>
<td>{food-1}</td>
<td>2123</td>
<td>{bocca-1}</td>
<td>157</td>
</tr>
<tr>
<td>CONTAINER</td>
<td>{container-1}</td>
<td>567</td>
<td>{bagna-1}</td>
<td>26</td>
</tr>
<tr>
<td>COVERING</td>
<td>{covering-4}</td>
<td>1024</td>
<td>{involucro-1}</td>
<td>40</td>
</tr>
<tr>
<td></td>
<td>{covering-5}</td>
<td></td>
<td>{copertura-2}</td>
<td></td>
</tr>
<tr>
<td>GARMENT</td>
<td>{wear-1}</td>
<td>277</td>
<td>{indumento-1}</td>
<td>156</td>
</tr>
<tr>
<td>PLACE</td>
<td>{location 1}</td>
<td>1881</td>
<td>{lugano-1}</td>
<td>54</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>{lugano-2}</td>
<td></td>
</tr>
<tr>
<td>FURNITURE</td>
<td>{furniture-1}</td>
<td>174</td>
<td>{meubelstuk-1}</td>
<td>75</td>
</tr>
<tr>
<td>PLANT</td>
<td>{plant-1}</td>
<td>802</td>
<td>{pianta-1}</td>
<td>474</td>
</tr>
<tr>
<td></td>
<td>{first 3 levels}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ANIMAL</td>
<td>{animal-1}</td>
<td>2017</td>
<td>{animale-1}</td>
<td>563</td>
</tr>
<tr>
<td>ARTIST</td>
<td>{artist-1}</td>
<td>71</td>
<td>{artista-1}</td>
<td>91</td>
</tr>
<tr>
<td>WORKER</td>
<td>{worker-2}</td>
<td>675</td>
<td>{lavoratore-1}</td>
<td>552</td>
</tr>
<tr>
<td>INSTRUMENT</td>
<td>{instrument 2}</td>
<td>509</td>
<td>{strumento-1}</td>
<td>867</td>
</tr>
<tr>
<td>VEHICLE</td>
<td>{vehicle 1}</td>
<td>410</td>
<td>{veicolo-1}</td>
<td>172</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>{transporte-5}</td>
<td></td>
</tr>
<tr>
<td>FEELINGS</td>
<td>{feeling-1}</td>
<td>448</td>
<td>{sentimiento-1}</td>
<td>178</td>
</tr>
<tr>
<td></td>
<td>{experience-6}</td>
<td></td>
<td>{percipire-1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>{feel-7}</td>
<td></td>
<td>{provare-7}</td>
<td></td>
</tr>
<tr>
<td>PHENOMENA</td>
<td>{phenomenon-1}</td>
<td>1012</td>
<td>{fenomeno-1}</td>
<td>100</td>
</tr>
<tr>
<td>COOKING</td>
<td>{cook-1}</td>
<td>57</td>
<td>{cuocere-1}</td>
<td>24</td>
</tr>
<tr>
<td></td>
<td>{cook-2}</td>
<td></td>
<td>{cuocere-2}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>{cook-3}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{cook-4}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{cooking-1}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MOVEMENT</td>
<td>{motion-1}</td>
<td>1891</td>
<td>{muoversi-1}</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td>{motion-2}</td>
<td></td>
<td>{muovere-1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>{motion-5}</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SOUNDS</td>
<td>{sound-13}</td>
<td>271</td>
<td>{sonido-2}</td>
<td>45</td>
</tr>
<tr>
<td></td>
<td>{sound-5}</td>
<td></td>
<td>{sonar-3}</td>
<td></td>
</tr>
<tr>
<td></td>
<td>{utter-3}</td>
<td></td>
<td>{emitir sonidos-1}</td>
<td></td>
</tr>
<tr>
<td>KNOWLEDGE</td>
<td>??</td>
<td>??</td>
<td>{conocencia-3}</td>
<td>223</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>{disciplina-1}</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>{conoscere-1}</td>
<td></td>
</tr>
</tbody>
</table>
For each cluster or for major hyperonyms within a cluster, the following data is given:

1. the number of hyponyms in each wordnet linked to the given hyperonyms and give an overview of the equivalence relations per equivalence type.
2. the concepts that occur in both the Source and Reference wordnets but have different classifications across the wordnets.
3. the concepts that cannot be projected from the Reference wordnets to the Source wordnet.

In the next table the results are listed for each cluster. The first column gives the number of (sub)-hyponyms in each field (WM). The second column for each wordnet gives the number of ILI-records that is linked to the (sub)-hyponyms or descendants. Note that there can be multiple ILIs for a single synset, that some synsets have no ILI and that different synsets may share the same ILI. There is thus a many-to-many mapping between the ILIs and the synsets and there can be less or more ILIs than descendants. Note that we omitted the number of ILIs for WordNet1.5 because it is currently the same. The third column gives the number of WMs that intersected (\(\cap\)) with the Source wordnet WMs for each semantic field. The source set is marked in the table. The fourth table then represents the percentage of this intersection of the total number of WMs in the source wordnet. The last column in the table gives the number of near-matches (NM) that have been recovered for each field.

Table 29: Projections and Intersections of comparing the First Subset

<table>
<thead>
<tr>
<th>Field</th>
<th>WordNet1.5 WM</th>
<th>Dutch Wordnet WM</th>
<th>Spanish Wordnet WM</th>
<th>Italian Wordnet WM</th>
<th>N M</th>
</tr>
</thead>
<tbody>
<tr>
<td>building</td>
<td>1220</td>
<td>170</td>
<td>48%</td>
<td>351</td>
<td>223</td>
</tr>
<tr>
<td>comestible</td>
<td>2156</td>
<td>103</td>
<td>54%</td>
<td>192</td>
<td>154</td>
</tr>
<tr>
<td>container</td>
<td>567</td>
<td>15</td>
<td>58%</td>
<td>26</td>
<td>31</td>
</tr>
<tr>
<td>covering</td>
<td>1024</td>
<td>18</td>
<td>67%</td>
<td>27</td>
<td>43</td>
</tr>
<tr>
<td>garment</td>
<td>277</td>
<td>127</td>
<td>59%</td>
<td>23</td>
<td>36</td>
</tr>
<tr>
<td>place</td>
<td>1881</td>
<td>373</td>
<td>100%</td>
<td>533</td>
<td>425</td>
</tr>
<tr>
<td>furniture</td>
<td>174</td>
<td>65</td>
<td>100%</td>
<td>11</td>
<td>15</td>
</tr>
<tr>
<td>plant</td>
<td>802</td>
<td>132</td>
<td>28%</td>
<td>28</td>
<td>40</td>
</tr>
<tr>
<td>animal</td>
<td>2017</td>
<td>311</td>
<td>55%</td>
<td>26</td>
<td>43</td>
</tr>
<tr>
<td>artist</td>
<td>71</td>
<td>3</td>
<td>3%</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>worker</td>
<td>675</td>
<td>287</td>
<td>52%</td>
<td>9</td>
<td>146</td>
</tr>
<tr>
<td>instrument</td>
<td>509</td>
<td>215</td>
<td>24%</td>
<td>437</td>
<td>266</td>
</tr>
<tr>
<td>vehicle</td>
<td>410</td>
<td>106</td>
<td>62%</td>
<td>21</td>
<td>35</td>
</tr>
<tr>
<td>feelings</td>
<td>456</td>
<td>25</td>
<td>28%</td>
<td>89</td>
<td>139</td>
</tr>
<tr>
<td>phenomena</td>
<td>1020</td>
<td>16</td>
<td>4%</td>
<td>356</td>
<td>241</td>
</tr>
<tr>
<td>cooking</td>
<td>57</td>
<td>20</td>
<td>100</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>sounds</td>
<td>271</td>
<td>139</td>
<td>100</td>
<td>22</td>
<td>33</td>
</tr>
<tr>
<td>movement</td>
<td>1891</td>
<td>95</td>
<td>64%</td>
<td>1313</td>
<td>1304</td>
</tr>
<tr>
<td>knowledge(^7)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>53</td>
<td>69</td>
</tr>
</tbody>
</table>

\(^7\) The comparison for the field knowledge could not be completed due to a technical problem.
The percentage of intersection directly reflect compatibility of wordnets. Extreme high intersection is found between WordNet1.5 and the Spanish wordnet, as can be expected given their methodology: PLACE, FURNITURE, COOKING, SOUND. However, many other projections still give a reasonable result around 50%, given the fact that we have not reached full coverage. If we look at intersections below 10% we first of all see that the Dutch wordnet lacks coverage (FURNITURE, PLANT, VEHICLE, COOKING, KNOWLEDGE) and that the Italian wordnet lacks equivalence relations (BUILDING, GARMENT, PLACE, FURNITURE). Both conclusions have already been suggested by previous data. A new fact is that PHENOMENA projected from WordNet1.5, ES and IT to Dutch, gives an extremely low intersection but still represents a large cluster in Dutch. This may point to a significant difference in classification in the Dutch wordnet. The following alternative classifications have been found for concepts which are PHENOMENA in the Reference wordnets (WordNet1.5, the Spanish wordnet and the Italian wordnet) but not in the Dutch wordnet:

- process/ change/ condition proces-2; verandering-1; gesteldheid-1 (all more general)
- systems: systeem (mechanisme)
- weather: weersgesteldheid (weather condition)
- power/force: energie-2 -> kracht-6 -> vermogen-; krachtveld
- possibilities: mogelijkheid
- diseases: ziekte-1

The alternative classifications show that there are many possibilities to describe a situation, which are not incompatible.

By further inspecting classification differences and the kind of equivalence mapping of all the fields we have come to the following conclusions:

1. Most mistakes are due to wrong translations, only a few are due to wrong classifications.
2. Alternative classifications occur quite regularly:
   - constructions: movable constructions; parts of buildings; institutions
   - comestibles: products such as fruits, grain, corn, seeds; drinks; parts
   - containers: object
   - covering: garments; parts of garments
   - feelings: stimulus (cause to feel like); more general experiences; attitudes; abilities
   - phenomena: process/ change/ condition; systems; weather conditions; power/force; possibilities; diseases
   - furniture: artifact or object
   - places: imaginary places; geographic terms; facility/installation (e.g. sports fields); containers
   - plants: microorganism; vegetables
   - sounds: communicate; breathe
   - cooking: creation, change
   - movement: sport; natural phenomena
3. There are quite a few cases of regular polysemy (e.g constructions and installations or facilities) which can be resolved by conflating word meanings in the ILI.
4. ES only has eq_synonym relations, while NL and IT also have other types of equivalences.
5. ES correspondence to ILI is practically one-to-one, while NL tends to have more ILIs than synsets; and IT less ILIs than synsets.
6. ES tends to have more synsets than the others —at this stage of the project --- covering most of the synsets projected from other the WNs.
3.2 Overall comparison of Subset1

3.2.1 Introduction

The main objective of the overall comparison is to measure the degree of coverage and intersection 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.

For this comparison each site (NL, IT, SP) has generated two sets (one for nouns and one for verbs) of hyponymy chains. For example, the next list of Dutch senses is generated for "opstijgen" (take off) by recursively taking all the hyperonyms:

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

To be able to compare these chains, each word sense in the chain has been replaced by eq_synonym and eq_near_synonym relations. When we reverse this chain (from top to bottom) we get the following result:

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

This means that the Dutch equivalent to ILI record number 00064108-v has as hyponym (the equivalent to ILI record number v 01046072-v) and this one has as hyponym (the equivalent to ILI record number 01046072-v), etc. Note that multiple translations lead to different chains.

In some cases (all of them for Dutch and Italian) an ILI chain contains nodes that have not been linked to WN1.5 equivalents. In these cases the original word instead of the ILI record number was used to identify the node. We then derived two statistics (when the differences are relevant) for chains with and without the untranslated nodes.

Two kinds of measurements have been developed: sense-based (synset or ILI) and chain based. Furthermore, the chain-based measurements have been divided into node-coverage and edge-coverage:

- 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 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
and
4--5

L2

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. There are, however, two subchains of length 3, one for each language, and 2 subchains of length 2, one for each language too.

Both measurements are important and can be used in different way. Of course edge-coverage is more 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. 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 languages

What must be done is:

1) assuring a high level of overlap in node-coverage
2) check if a possible lower edge-coverage is due to either b) incompleteness (to be corrected) or c) real differences.

Complete overlapping of chains (either at edge or node level) is difficult to achieve in the case of huge differences in the size of the wordnets to be compared (e.g. the nouns in the Spanish wordnet, which is the largest subset, still hardly covers 30% of the nouns in WN1.5). We have therefore developed two other kind of measurements that are more useful for comparison: subsequences and subsequences with gaps:

- Subsequences are simply chains of nodes/edges that exactly match a fragment of another chain. Subsequences can be classified according to their length.
- Subsequences with n gaps are chains of nodes/edges that match a fragment of another chain but failing to match n nodes of edges.

For example:

- **Node subsequence of length 2:**
  Sequence:
  00002728 00004865 05839075 06193747
  Subsequence:
  00004865 05839075

- **Edge subsequence of length 2:**
  Sequence:
  00002728 00004865 05839075 06193747
  Subsequence:
  00004865 05839075 06193747

- **Node subsequence of length 3 with 1 gap:**
  Sequence:
  00002728 00004865 05839075 06193747
  Subsequence:
  00004865 06193747

- **Edge subsequence of length 4 with 2 gap:**
  Sequence:
  00002728 00004865 05839075 06193747 01137195
  Subsequence:
  00002728 00004865 06193747 01137195

Subsequences with 1 and 2 gaps are reported here. Although other cases can be computed in an easy way, their usefulness is clearly lower.

The procedure we have developed in order to extract the statistics consists of four steps:

1. One of the WNs is taken as 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 skeleton. 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 (see also Appendix II). When using the verbose mode, not only the number but also the actual occurrences of the overlapping cases are extracted. Some of these instances are presented as examples for each kind of query. Normal mode is used here for presenting the results and extracting some conclusions from them in order to improve the overlapping of the different wordnets. Verbose mode will be used in a 2nd step for selecting ILI nodes or edges partially uncovered and guiding the extension of subset1.

Each site has generated an ASCII file containing the ILI-chains for all synsets in their Subset1. The following statistics can be extracted:

1. Frequencies and ratios of chains / length /language
2. Frequencies and ratios of ILI records / language
3. chains completely overlapping for pairs of languages (6), triples (3) and all 4 languages (if any). (occurrences, frequencies and ratios)
4. The same as 3) but for sub-chains of length N, for any value of N.
5. The same as 4) but allowing a maximum of M gaps (non overlapping elements within the subchain). These gaps could be contiguous or not.

In appendix II, we have listed the programs and software utilities that have been used to extract the data.

### 3.2.2 Evaluation of individual wordnets

The next results are taken from an analysis of individual chain sets. We performed the whole process for those ILI-records having WN1.5 equivalents. The figures are computed prior to any process. The content of the different columns is the following:

ILI nodes: number of different ILI nodes appearing in the sets (only ILI numbers having a WN1.5 translation have been taken into account because the comparison in the other cases is at this moment impossible)

- tops: nodes with no hyponym link
- leaves: nodes with no hypernym link
- internal nodes: nodes not being tops or leaves
- edges: number of edges appearing in the sets
- chains: number of chains appearing in the sets

#### Table 30: ILI chains for nouns

<table>
<thead>
<tr>
<th></th>
<th>ILI nodes</th>
<th>Tops</th>
<th>Leaves</th>
<th>Internal Nodes</th>
<th>EDGES</th>
<th>CHAINS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>18577</td>
<td>11</td>
<td>14208</td>
<td>4358</td>
<td>18885</td>
<td>16784</td>
</tr>
<tr>
<td>IT</td>
<td>1608</td>
<td>18</td>
<td>1446</td>
<td>161</td>
<td>2390</td>
<td>5083</td>
</tr>
<tr>
<td>NL</td>
<td>5098</td>
<td>14</td>
<td>5091</td>
<td>1510</td>
<td>6124</td>
<td>14673</td>
</tr>
<tr>
<td>WN15</td>
<td>60557</td>
<td>11</td>
<td>47110</td>
<td>13436</td>
<td>61123</td>
<td>53467</td>
</tr>
</tbody>
</table>

#### Table 31: ILI chains for verbs

<table>
<thead>
<tr>
<th></th>
<th>ILI nodes</th>
<th>Tops</th>
<th>Leaves</th>
<th>Internal Nodes</th>
<th>EDGES</th>
<th>CHAINS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>3218</td>
<td>368</td>
<td>2377</td>
<td>673</td>
<td>2863</td>
<td>2393</td>
</tr>
<tr>
<td>IT</td>
<td>541</td>
<td>39</td>
<td>447</td>
<td>60</td>
<td>790</td>
<td>849</td>
</tr>
<tr>
<td>NL</td>
<td>2142</td>
<td>13</td>
<td>2135</td>
<td>807</td>
<td>3351</td>
<td>5136</td>
</tr>
<tr>
<td>WN15</td>
<td>11363</td>
<td>573</td>
<td>8446</td>
<td>2580</td>
<td>10816</td>
<td>8486</td>
</tr>
</tbody>
</table>

---

8 in some cases an ILI node appears both as leaf and as internal node, so Tops, Leaves and Internal nodes are not disjoint sets. So the result of adding the Tops, Leaves and Internal nodes could be greater than the number of ILI nodes.
Table 32: ILI chains for nouns and verbs

<table>
<thead>
<tr>
<th>ILI nodes</th>
<th>ILL nodes</th>
<th>Tops</th>
<th>Leaves</th>
<th>Internal Nodes</th>
<th>EDGES</th>
<th>CHAINS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>21795</td>
<td>379</td>
<td>16585</td>
<td>5031</td>
<td>21748</td>
<td>19177</td>
</tr>
<tr>
<td>IT</td>
<td>2149</td>
<td>57</td>
<td>1893</td>
<td>221</td>
<td>3180</td>
<td>5932</td>
</tr>
<tr>
<td>NL</td>
<td>7240</td>
<td>27</td>
<td>7226</td>
<td>2317</td>
<td>9475</td>
<td>19809</td>
</tr>
<tr>
<td>WN15</td>
<td>71920</td>
<td>584</td>
<td>55556</td>
<td>16016</td>
<td>71939</td>
<td>61953</td>
</tr>
</tbody>
</table>

The next tables show the same data as before but9:

1) Removing chains including inconsistent information, e.g. for nouns, those chains with verbs or adjectives.
2) Removing garbage (null lines, etc.).
3) Replacing repeated ILI-nodes in a chain for only one occurrence.
4) Removing those chains included in another ones (the hypernym chains).
5) Removing the repeated chains.

Table 33: ILI chains for nouns

<table>
<thead>
<tr>
<th>ILI nodes</th>
<th>%WN</th>
<th>Tops</th>
<th>Leaves</th>
<th>Internal Nodes</th>
<th>EDGES</th>
<th>%WN</th>
<th>CHAINS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>30.68</td>
<td>11</td>
<td>14208</td>
<td>4358</td>
<td>18885</td>
<td>30.90</td>
<td>16784</td>
</tr>
<tr>
<td>IT</td>
<td>-</td>
<td>17</td>
<td>1444</td>
<td>161</td>
<td>2390</td>
<td>-</td>
<td>4671</td>
</tr>
<tr>
<td>IT2(*)</td>
<td>1.84</td>
<td>17</td>
<td>977</td>
<td>135</td>
<td>1664</td>
<td>2.72</td>
<td>3197</td>
</tr>
<tr>
<td>NL</td>
<td>-</td>
<td>35</td>
<td>1419</td>
<td>5883</td>
<td>-</td>
<td>14661</td>
<td></td>
</tr>
<tr>
<td>NL2(**)</td>
<td>6.28</td>
<td>13</td>
<td>3807</td>
<td>1129</td>
<td>4491</td>
<td>7.35</td>
<td>14661</td>
</tr>
<tr>
<td>WN15</td>
<td>100</td>
<td>11</td>
<td>47110</td>
<td>13436</td>
<td>61123</td>
<td>100</td>
<td>53467</td>
</tr>
</tbody>
</table>

(*) IT2: is Italian WordNet without exclusive Italian ILI Records.
(**) NL2: is Dutch WordNet without exclusive Dutch ILI Records.

Table 34: clean ILI chains for verbs

<table>
<thead>
<tr>
<th>ILI nodes</th>
<th>%WN</th>
<th>Tops</th>
<th>Leaves</th>
<th>Internal Nodes</th>
<th>EDGES</th>
<th>%WN</th>
<th>CHAINS</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>28.32</td>
<td>368</td>
<td>2377</td>
<td>673</td>
<td>2863</td>
<td>26.47</td>
<td>2393</td>
</tr>
<tr>
<td>IT</td>
<td>-</td>
<td>39</td>
<td>445</td>
<td>60</td>
<td>790</td>
<td>-</td>
<td>783</td>
</tr>
<tr>
<td>IT2(*)</td>
<td>3.85</td>
<td>39</td>
<td>359</td>
<td>41</td>
<td>634</td>
<td>5.86</td>
<td>629</td>
</tr>
<tr>
<td>NL</td>
<td>-</td>
<td>12</td>
<td>2080</td>
<td>743</td>
<td>3088</td>
<td>-</td>
<td>4968</td>
</tr>
<tr>
<td>NL2(**)</td>
<td>15.67</td>
<td>12</td>
<td>1776</td>
<td>673</td>
<td>2690</td>
<td>24.87</td>
<td>4968</td>
</tr>
<tr>
<td>WN15</td>
<td>100</td>
<td>573</td>
<td>8446</td>
<td>2580</td>
<td>10816</td>
<td>100</td>
<td>8486</td>
</tr>
</tbody>
</table>

(*) IT2: is Italian WordNet without exclusive Italian ILI Records.
(**) NL2: is Dutch WordNet without exclusive Dutch ILI Records.

Some conclusions can be extracted from these results:

1) There is a (relatively) high number of Italian and Dutch nodes without WN1.5 equivalent (31% of Italian nouns, 25% of Dutch nouns, 19% of Italian verbs, 14% of Dutch verbs). Without equivalents for these nodes, complete overlapping is not possible, not only at the node level but also, to a great extent, for the chains including such nodes.
2) There is (in general) a slightly better coverage in terms of edges than in term of nodes. This is a positive result because it seems to indicate that, when present, nodes are connected in a consistent way.

---

9For Italian and Dutch we present the data in two forms: raw data and chains where nodes without WN1.5 equivalent are dropped out.
The next two tables present the number and % of noun and verb chains classified by length for each language.

**Table 35: Frequencies and ratios of noun chains / length /language**

<table>
<thead>
<tr>
<th>WN frequency</th>
<th>NL frequency</th>
<th>IT frequency</th>
<th>ES frequency</th>
<th>%</th>
<th>%</th>
<th>%</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>33</td>
<td>0.06</td>
<td>21</td>
<td>1.17</td>
<td>642</td>
<td>13.63</td>
<td>86</td>
</tr>
<tr>
<td>3</td>
<td>521</td>
<td>0.97</td>
<td>509</td>
<td>4.14</td>
<td>1662</td>
<td>35.29</td>
<td>761</td>
</tr>
<tr>
<td>4</td>
<td>2220</td>
<td>4.15</td>
<td>1423</td>
<td>11.59</td>
<td>445</td>
<td>9.45</td>
<td>1678</td>
</tr>
<tr>
<td>5</td>
<td>5664</td>
<td>10.59</td>
<td>2306</td>
<td>18.78</td>
<td>925</td>
<td>19.64</td>
<td>3088</td>
</tr>
<tr>
<td>6</td>
<td>12730</td>
<td>23.81</td>
<td>2656</td>
<td>21.63</td>
<td>543</td>
<td>11.53</td>
<td>4010</td>
</tr>
<tr>
<td>7</td>
<td>11741</td>
<td>21.96</td>
<td>2708</td>
<td>22.05</td>
<td>304</td>
<td>6.46</td>
<td>3565</td>
</tr>
<tr>
<td>8</td>
<td>8737</td>
<td>16.34</td>
<td>1651</td>
<td>13.44</td>
<td>188</td>
<td>3.99</td>
<td>2204</td>
</tr>
<tr>
<td>9</td>
<td>5940</td>
<td>11.11</td>
<td>708</td>
<td>5.76</td>
<td>0</td>
<td>0</td>
<td>849</td>
</tr>
<tr>
<td>10</td>
<td>3305</td>
<td>6.18</td>
<td>208</td>
<td>1.69</td>
<td>0</td>
<td>0</td>
<td>344</td>
</tr>
<tr>
<td>11</td>
<td>1400</td>
<td>2.62</td>
<td>56</td>
<td>0.46</td>
<td>0</td>
<td>0</td>
<td>121</td>
</tr>
<tr>
<td>12</td>
<td>517</td>
<td>0.97</td>
<td>24</td>
<td>0.20</td>
<td>0</td>
<td>0</td>
<td>27</td>
</tr>
<tr>
<td>13</td>
<td>364</td>
<td>0.68</td>
<td>5</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
<td>11</td>
</tr>
<tr>
<td>14</td>
<td>213</td>
<td>0.4</td>
<td>2</td>
<td>0.02</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>75</td>
<td>0.14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>7</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>53467</td>
<td>100</td>
<td>12282</td>
<td>100</td>
<td>4709</td>
<td>100</td>
<td>16744</td>
</tr>
</tbody>
</table>

**Average**

<table>
<thead>
<tr>
<th>WN frequency</th>
<th>NL frequency</th>
<th>IT frequency</th>
<th>ES frequency</th>
<th>%</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>236</td>
<td>2.79</td>
<td>3</td>
<td>0.01</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>1837</td>
<td>27.72</td>
<td>102</td>
<td>0.43</td>
<td>451</td>
<td>42.42</td>
</tr>
<tr>
<td>2530</td>
<td>29.92</td>
<td>260</td>
<td>1.09</td>
<td>90</td>
<td>8.47</td>
</tr>
<tr>
<td>1959</td>
<td>23.17</td>
<td>838</td>
<td>3.52</td>
<td>125</td>
<td>11.76</td>
</tr>
<tr>
<td>1029</td>
<td>12.17</td>
<td>1328</td>
<td>5.58</td>
<td>201</td>
<td>18.91</td>
</tr>
<tr>
<td>462</td>
<td>5.46</td>
<td>1977</td>
<td>8.31</td>
<td>142</td>
<td>13.36</td>
</tr>
<tr>
<td>250</td>
<td>2.96</td>
<td>2284</td>
<td>9.60</td>
<td>42</td>
<td>3.95</td>
</tr>
<tr>
<td>109</td>
<td>1.29</td>
<td>2520</td>
<td>10.59</td>
<td>12</td>
<td>1.13</td>
</tr>
<tr>
<td>32</td>
<td>0.38</td>
<td>2562</td>
<td>10.77</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>0.12</td>
<td>2769</td>
<td>11.64</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>0.02</td>
<td>2512</td>
<td>10.36</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>2152</td>
<td>9.05</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1571</td>
<td>6.60</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1080</td>
<td>4.54</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>779</td>
<td>3.27</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>532</td>
<td>2.24</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>277</td>
<td>1.16</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>132</td>
<td>0.55</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>68</td>
<td>0.29</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>31</td>
<td>0.13</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0.04</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.00</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>8456</td>
<td>100</td>
<td>23787</td>
<td>100</td>
<td>1063</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>3.58</td>
<td>9.59</td>
<td>3.69</td>
<td>3.01</td>
<td></td>
</tr>
</tbody>
</table>

**Table 36: Frequencies and ratios of verb chains / length /language**

<table>
<thead>
<tr>
<th>WN frequency</th>
<th>NL frequency</th>
<th>IT frequency</th>
<th>ES frequency</th>
<th>%</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>236</td>
<td>2.79</td>
<td>3</td>
<td>0.01</td>
<td>0</td>
</tr>
<tr>
<td>2</td>
<td>1837</td>
<td>27.72</td>
<td>102</td>
<td>0.43</td>
<td>451</td>
</tr>
<tr>
<td>3</td>
<td>2530</td>
<td>29.92</td>
<td>260</td>
<td>1.09</td>
<td>90</td>
</tr>
<tr>
<td>4</td>
<td>1959</td>
<td>23.17</td>
<td>838</td>
<td>3.52</td>
<td>125</td>
</tr>
<tr>
<td>5</td>
<td>1029</td>
<td>12.17</td>
<td>1328</td>
<td>5.58</td>
<td>201</td>
</tr>
<tr>
<td>6</td>
<td>462</td>
<td>5.46</td>
<td>1977</td>
<td>8.31</td>
<td>142</td>
</tr>
<tr>
<td>7</td>
<td>250</td>
<td>2.96</td>
<td>2284</td>
<td>9.60</td>
<td>42</td>
</tr>
<tr>
<td>8</td>
<td>109</td>
<td>1.29</td>
<td>2520</td>
<td>10.59</td>
<td>12</td>
</tr>
<tr>
<td>9</td>
<td>32</td>
<td>0.38</td>
<td>2562</td>
<td>10.77</td>
<td>0</td>
</tr>
<tr>
<td>10</td>
<td>10</td>
<td>0.12</td>
<td>2769</td>
<td>11.64</td>
<td>0</td>
</tr>
<tr>
<td>11</td>
<td>2</td>
<td>0.02</td>
<td>2512</td>
<td>10.36</td>
<td>0</td>
</tr>
<tr>
<td>12</td>
<td>0</td>
<td>0</td>
<td>2152</td>
<td>9.05</td>
<td>0</td>
</tr>
<tr>
<td>13</td>
<td>0</td>
<td>0</td>
<td>1571</td>
<td>6.60</td>
<td>0</td>
</tr>
<tr>
<td>14</td>
<td>0</td>
<td>0</td>
<td>1080</td>
<td>4.54</td>
<td>0</td>
</tr>
<tr>
<td>15</td>
<td>0</td>
<td>0</td>
<td>779</td>
<td>3.27</td>
<td>0</td>
</tr>
<tr>
<td>16</td>
<td>0</td>
<td>0</td>
<td>532</td>
<td>2.24</td>
<td>0</td>
</tr>
<tr>
<td>17</td>
<td>0</td>
<td>0</td>
<td>277</td>
<td>1.16</td>
<td>0</td>
</tr>
<tr>
<td>18</td>
<td>0</td>
<td>0</td>
<td>132</td>
<td>0.55</td>
<td>0</td>
</tr>
<tr>
<td>19</td>
<td>0</td>
<td>0</td>
<td>68</td>
<td>0.29</td>
<td>0</td>
</tr>
<tr>
<td>20</td>
<td>0</td>
<td>0</td>
<td>31</td>
<td>0.13</td>
<td>0</td>
</tr>
<tr>
<td>21</td>
<td>0</td>
<td>0</td>
<td>9</td>
<td>0.04</td>
<td>0</td>
</tr>
<tr>
<td>22</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0.00</td>
<td>0</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td>8456</td>
<td>100</td>
<td>23787</td>
<td>100</td>
<td>1063</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td>3.58</td>
<td>9.59</td>
<td>3.69</td>
<td>3.01</td>
<td></td>
</tr>
</tbody>
</table>

Conclusions extracted from these tables:

1) the distribution in the case of nouns appears to be quite nice, at least it follows quite closely WN1.5 distribution.

The differences in length (we can use the average length as measure) can be explained by the lower coverage (e.g. 7.19 in the case of WN1.5 vs 6.22 in the case of Spanish or Dutch).
2) In the case of Italian the average length is lower (4.15). The lower size of Italian subset1 can account partially for this difference. Another reason could be the concentration of Italian nominal nodes in the higher levels of the hierarchy or simply the lack of hyponymy/hypernym relations. The fact that with a nominal coverage significantly lower than the Spanish or Dutch (1608 ILI nodes vs 18577 or 5090) the number of tops is higher (17 vs 11 or 13) seems to point to this later explanation.

3) In the case of verbs the figures for Italian and Spanish wordnets are very close to WN1.5. This is not the case for the Dutch wordnet, which has long chains. A careful examination of some of these pathological chains shows that they come from wrong translation equivalences to WN1.5. If several (not many) of these equivalences appear in the higher levels of the hierarchy they have a multiplicative effect on the number of chains and the results appearing in the table are absolutely useless. The solution is to analyze carefully the most frequent long chains for checking the equivalence relations involved in them. There is no need to analyze all the long chains (i.e. those with length greater than 10) but to find the most frequent subsequences appearing in those long chains.

3.2.3 Global evaluation

The next three tables account for the coverage of the individual wordnets (NL, IT, SP), pairs (NL-IT, NL-SP, IT-SP) and full intersection (NL-IT-SP) against 1) WN1.5 and 2) the union NL-IT-SP

Table 37: Coverage of noun ILI records

<table>
<thead>
<tr>
<th>Total frequency</th>
<th>(60557)</th>
<th>(21828)</th>
<th>% WN</th>
<th>% ∪(IT, NL, ES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>18577</td>
<td>30.68</td>
<td>85.11</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>1608</td>
<td>2.66</td>
<td>7.37</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>5090</td>
<td>8.41</td>
<td>23.32</td>
<td></td>
</tr>
<tr>
<td>∩ (ES, IT)</td>
<td>853</td>
<td>1.41</td>
<td>3.91</td>
<td></td>
</tr>
<tr>
<td>∩ (ES, NL)</td>
<td>2539</td>
<td>4.19</td>
<td>11.63</td>
<td></td>
</tr>
<tr>
<td>∩ (IT, NL)</td>
<td>389</td>
<td>0.64</td>
<td>1.78</td>
<td></td>
</tr>
<tr>
<td>∩ (ES, IT, NL)</td>
<td>334</td>
<td>0.55</td>
<td>1.53</td>
<td></td>
</tr>
</tbody>
</table>

Table 38: Coverage of verb ILI records

<table>
<thead>
<tr>
<th>Total frequency</th>
<th>(11363)</th>
<th>(4592)</th>
<th>% WN</th>
<th>% ∪(IT, NL, ES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>3224</td>
<td>28.37</td>
<td>70.21</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>541</td>
<td>4.76</td>
<td>11.78</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>2085</td>
<td>18.35</td>
<td>45.41</td>
<td></td>
</tr>
<tr>
<td>∩ (ES, IT)</td>
<td>307</td>
<td>2.70</td>
<td>6.69</td>
<td></td>
</tr>
<tr>
<td>∩ (ES, NL)</td>
<td>918</td>
<td>8.08</td>
<td>19.99</td>
<td></td>
</tr>
<tr>
<td>∩ (IT, NL)</td>
<td>198</td>
<td>1.74</td>
<td>4.31</td>
<td></td>
</tr>
<tr>
<td>∩ (ES, IT, NL)</td>
<td>165</td>
<td>1.45</td>
<td>3.59</td>
<td></td>
</tr>
</tbody>
</table>

Table 39: Coverage of ILI records (total)

<table>
<thead>
<tr>
<th>Total frequency</th>
<th>(71920)</th>
<th>(26420)</th>
<th>% WN</th>
<th>% ∪(IT, NL, ES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>21801</td>
<td>30.31</td>
<td>82.52</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>2149</td>
<td>2.99</td>
<td>8.13</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>7175</td>
<td>9.98</td>
<td>21.16</td>
<td></td>
</tr>
<tr>
<td>∩ (ES, IT)</td>
<td>1160</td>
<td>1.61</td>
<td>4.39</td>
<td></td>
</tr>
<tr>
<td>∩ (ES, NL)</td>
<td>3457</td>
<td>4.81</td>
<td>13.08</td>
<td></td>
</tr>
<tr>
<td>∩ (IT, NL)</td>
<td>587</td>
<td>0.82</td>
<td>2.22</td>
<td></td>
</tr>
<tr>
<td>∩ (ES, IT, NL)</td>
<td>499</td>
<td>0.69</td>
<td>1.89</td>
<td></td>
</tr>
</tbody>
</table>

It is difficult to extract conclusions from these three tables. An obvious objective is to increase the intersection between languages but it is difficult to achieve this objective in the same way we performed the construction of the Base Concepts. For one thing, we see that the union of all the ILIs (26420 nodes) almost represents the maximal set of

10 This is the result of the strategy to include the best two translations that have been generated by the automatic matching program. Since most verbs get several solutions a relatively high percentage of mistakes is generated.
synsets we aim at (30,000 synsets). Furthermore, not all these synsets are lexicalized in all the languages. However, we think that an improvement of the intersection can be obtained in an indirect way by filling gaps (or by extending subchains). These possibilities will be presented and discussed later.

The next ten tables account for the coverage of complete chains (at node and edge level) for nouns and verbs, projected over the different WNs. In the case of WN1.5, all the other WNs have been projected over while in the other cases WN1.5 has not been taken into account.

**Table 40: Coverage of complete noun chains projected over WN1.5 structure**

<table>
<thead>
<tr>
<th></th>
<th>nodes (53467)</th>
<th>edges (53467)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>frequency %</td>
<td>frequency %</td>
</tr>
<tr>
<td>ES</td>
<td>7539</td>
<td>14.10</td>
</tr>
<tr>
<td>IT</td>
<td>3</td>
<td>0.01</td>
</tr>
<tr>
<td>NL</td>
<td>288</td>
<td>0.54</td>
</tr>
<tr>
<td>∩(ES, IT)</td>
<td>2</td>
<td>0.00</td>
</tr>
<tr>
<td>∩(ES, NL)</td>
<td>150</td>
<td>0.28</td>
</tr>
<tr>
<td>∩(IT, NL)</td>
<td>1</td>
<td>0.00</td>
</tr>
<tr>
<td>∩(ES, IT, NL)</td>
<td>1</td>
<td>0.00</td>
</tr>
</tbody>
</table>

**Table 41: Coverage of complete verb chains projected over WN1.5 structure**

<table>
<thead>
<tr>
<th></th>
<th>nodes (8486)</th>
<th>edges (8486)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>frequency %</td>
<td>frequency %</td>
</tr>
<tr>
<td>ES</td>
<td>1235</td>
<td>14.55</td>
</tr>
<tr>
<td>IT</td>
<td>67</td>
<td>0.79</td>
</tr>
<tr>
<td>NL</td>
<td>413</td>
<td>4.87</td>
</tr>
<tr>
<td>∩(ES, IT)</td>
<td>33</td>
<td>0.39</td>
</tr>
<tr>
<td>∩(ES, NL)</td>
<td>133</td>
<td>1.57</td>
</tr>
<tr>
<td>∩(IT, NL)</td>
<td>15</td>
<td>0.18</td>
</tr>
<tr>
<td>∩(ES, IT, NL)</td>
<td>12</td>
<td>0.14</td>
</tr>
</tbody>
</table>

**Table 42: Coverage of complete noun chains projected over Dutch wordnet**

<table>
<thead>
<tr>
<th></th>
<th>nodes (12282)</th>
<th>edges (12282)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>frequency %</td>
<td>frequency %</td>
</tr>
<tr>
<td>ES</td>
<td>3006</td>
<td>24.47</td>
</tr>
<tr>
<td>IT</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>∩(ES, IT)</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Table 43: Coverage of complete verb chains projected over Dutch wordnet**

<table>
<thead>
<tr>
<th></th>
<th>nodes (23787)</th>
<th>edges (23787)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>frequency %</td>
<td>frequency %</td>
</tr>
<tr>
<td>ES</td>
<td>1172</td>
<td>4.93</td>
</tr>
<tr>
<td>IT</td>
<td>21</td>
<td>0.09</td>
</tr>
<tr>
<td>∩(ES, IT)</td>
<td>13</td>
<td>0.05</td>
</tr>
</tbody>
</table>

**Table 44: Coverage of complete noun chains projected over Italian wordnet**

<table>
<thead>
<tr>
<th></th>
<th>nodes (4709)</th>
<th>edges (4709)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>frequency %</td>
<td>frequency %</td>
</tr>
<tr>
<td>ES</td>
<td>2004</td>
<td>42.56</td>
</tr>
<tr>
<td>NL</td>
<td>238</td>
<td>5.05</td>
</tr>
<tr>
<td>∩(ES, NL)</td>
<td>177</td>
<td>3.76</td>
</tr>
</tbody>
</table>
The figures presented in the preceding tables are of rather limited use, since full coverage of the chains is not possible. The coverage in terms of complete chains is extremely low and the reason is the great differences in size between the different wordnets. Consider, for instance, the overlapping between WN1.5 and the Spanish wordnet. The ratio, when comparing nodes is for nouns 30.68%. When comparing full chains this figure drops to 14.10%. This is not necessarily a bad result. A possible interpretation could be that most of the coverage is concentrated in the highest levels of the hierarchy. This is confirmed by other evidence. Obviously better, and more useful, results will be obtained when dealing with incomplete sequences (both subsequences and sequences containing gaps). These cases will be considered in next.

The following four tables account for the overlapping of partial chains (node vs edge, noun vs verb) projected over WN1.5 structure, for different lengths of the chain

### Table 45: Coverage of complete verb chains projected over Italian wordnet

<table>
<thead>
<tr>
<th></th>
<th>nodes (1063)</th>
<th>edges (1063)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>205</td>
<td>39</td>
</tr>
<tr>
<td>NL</td>
<td>135</td>
<td>17</td>
</tr>
<tr>
<td>∩(ES, NL)</td>
<td>75</td>
<td>5</td>
</tr>
</tbody>
</table>

### Table 46: Coverage of complete noun chains projected over Spanish wordnet

<table>
<thead>
<tr>
<th></th>
<th>nodes (16744)</th>
<th>edges (16744)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL</td>
<td>380</td>
<td>8</td>
</tr>
<tr>
<td>IT</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>∩(NL, IT)</td>
<td>1</td>
<td>0</td>
</tr>
</tbody>
</table>

### Table 47: Coverage of complete verb chains projected over Spanish wordnet

<table>
<thead>
<tr>
<th></th>
<th>nodes (2397)</th>
<th>edges (2397)</th>
</tr>
</thead>
<tbody>
<tr>
<td>NL</td>
<td>297</td>
<td>89</td>
</tr>
<tr>
<td>IT</td>
<td>72</td>
<td>50</td>
</tr>
<tr>
<td>∩(NL, IT)</td>
<td>25</td>
<td>14</td>
</tr>
</tbody>
</table>

### Table 48: Coverage of partial noun chains of NODES projected over WN1.5 structure

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>NL</th>
<th>∩(ES, NL)</th>
<th>∩(ES, IT)</th>
<th>∩(IT, NL)</th>
<th>∩(ES, IT, NL)</th>
<th>WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>53467</td>
<td>36102</td>
<td>52504</td>
<td>52376</td>
<td>36080</td>
<td>30912</td>
<td>30909</td>
<td>53467</td>
</tr>
<tr>
<td>2</td>
<td>47792</td>
<td>23597</td>
<td>39616</td>
<td>38828</td>
<td>23531</td>
<td>16181</td>
<td>16151</td>
<td>53467</td>
</tr>
<tr>
<td>3</td>
<td>45744</td>
<td>15219</td>
<td>24004</td>
<td>23448</td>
<td>15189</td>
<td>6767</td>
<td>6756</td>
<td>53434</td>
</tr>
<tr>
<td>4</td>
<td>41747</td>
<td>7896</td>
<td>14200</td>
<td>14009</td>
<td>7844</td>
<td>2048</td>
<td>2001</td>
<td>52913</td>
</tr>
<tr>
<td>5</td>
<td>23930</td>
<td>3709</td>
<td>6146</td>
<td>5659</td>
<td>3627</td>
<td>809</td>
<td>780</td>
<td>50693</td>
</tr>
<tr>
<td>6</td>
<td>23350</td>
<td>948</td>
<td>2888</td>
<td>2535</td>
<td>751</td>
<td>402</td>
<td>393</td>
<td>45029</td>
</tr>
<tr>
<td>7</td>
<td>11774</td>
<td>333</td>
<td>1095</td>
<td>901</td>
<td>265</td>
<td>236</td>
<td>228</td>
<td>32299</td>
</tr>
<tr>
<td>8</td>
<td>5007</td>
<td>93</td>
<td>435</td>
<td>374</td>
<td>63</td>
<td>11</td>
<td>9</td>
<td>20558</td>
</tr>
<tr>
<td>9</td>
<td>1795</td>
<td>14</td>
<td>73</td>
<td>67</td>
<td>9</td>
<td>0</td>
<td>0</td>
<td>11821</td>
</tr>
<tr>
<td>10</td>
<td>664</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5881</td>
</tr>
<tr>
<td>11</td>
<td>137</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2576</td>
</tr>
<tr>
<td>12</td>
<td>36</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1176</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>659</td>
</tr>
</tbody>
</table>
Table 49: Coverage of partial noun chains of EDGES projected over WN1.5 structure

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>NL</th>
<th>(∩(ES, NL))</th>
<th>(∩(ES, IT))</th>
<th>(∩(IT, NL))</th>
<th>(∩(ES, IT, NL))</th>
<th>WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>47792</td>
<td>15185</td>
<td>29409</td>
<td>28928</td>
<td>15154</td>
<td>9168</td>
<td>9133</td>
<td>53467</td>
</tr>
<tr>
<td>2</td>
<td>45744</td>
<td>816</td>
<td>13647</td>
<td>13540</td>
<td>636</td>
<td>271</td>
<td>253</td>
<td>53434</td>
</tr>
<tr>
<td>3</td>
<td>41747</td>
<td>105</td>
<td>367</td>
<td>214</td>
<td>93</td>
<td>46</td>
<td>46</td>
<td>52913</td>
</tr>
<tr>
<td>4</td>
<td>23930</td>
<td>4</td>
<td>53</td>
<td>18</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>50693</td>
</tr>
<tr>
<td>5</td>
<td>23350</td>
<td>0</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>45029</td>
</tr>
<tr>
<td>6</td>
<td>11774</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>32299</td>
</tr>
<tr>
<td>7</td>
<td>5007</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20558</td>
</tr>
<tr>
<td>8</td>
<td>1795</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>11821</td>
</tr>
<tr>
<td>9</td>
<td>664</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>5881</td>
</tr>
<tr>
<td>10</td>
<td>137</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2576</td>
</tr>
<tr>
<td>11</td>
<td>36</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1176</td>
</tr>
<tr>
<td>12</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>659</td>
</tr>
</tbody>
</table>

Table 50: Coverage of partial VERB chains of NODES projected over WN1.5 structure

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>NL</th>
<th>(∩(ES, NL))</th>
<th>(∩(ES, IT))</th>
<th>(∩(IT, NL))</th>
<th>(∩(ES, IT, NL))</th>
<th>WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>7861</td>
<td>6299</td>
<td>7188</td>
<td>6720</td>
<td>6248</td>
<td>5423</td>
<td>5382</td>
<td>8486</td>
</tr>
<tr>
<td>2</td>
<td>5273</td>
<td>1601</td>
<td>3334</td>
<td>2563</td>
<td>1483</td>
<td>840</td>
<td>765</td>
<td>8250</td>
</tr>
<tr>
<td>3</td>
<td>2980</td>
<td>233</td>
<td>1244</td>
<td>799</td>
<td>208</td>
<td>46</td>
<td>43</td>
<td>6383</td>
</tr>
<tr>
<td>4</td>
<td>1184</td>
<td>37</td>
<td>234</td>
<td>112</td>
<td>37</td>
<td>0</td>
<td>0</td>
<td>3853</td>
</tr>
<tr>
<td>5</td>
<td>107</td>
<td>0</td>
<td>24</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1894</td>
</tr>
<tr>
<td>6</td>
<td>82</td>
<td>0</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>865</td>
</tr>
<tr>
<td>7</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>403</td>
</tr>
<tr>
<td>8</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>153</td>
</tr>
</tbody>
</table>

Table 51: Coverage of partial VERB chains of EDGES projected over WN1.5 structure

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>NL</th>
<th>(∩(ES, NL))</th>
<th>(∩(ES, IT))</th>
<th>(∩(IT, NL))</th>
<th>(∩(ES, IT, NL))</th>
<th>WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>5273</td>
<td>701</td>
<td>1406</td>
<td>1056</td>
<td>621</td>
<td>193</td>
<td>150</td>
<td>8250</td>
</tr>
<tr>
<td>2</td>
<td>2980</td>
<td>41</td>
<td>83</td>
<td>67</td>
<td>39</td>
<td>1</td>
<td>0</td>
<td>6383</td>
</tr>
<tr>
<td>3</td>
<td>1184</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3853</td>
</tr>
<tr>
<td>4</td>
<td>107</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1894</td>
</tr>
<tr>
<td>5</td>
<td>82</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>865</td>
</tr>
<tr>
<td>6</td>
<td>14</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>403</td>
</tr>
<tr>
<td>7</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>153</td>
</tr>
</tbody>
</table>

A short explanation is needed for interpreting these tables. In the case of node coverage a subsequence of length 1 corresponds to a simple node. As for WN1.5 all the sequences start with a top node, this means that for nouns every possible chain contains one of the 11 tops. In this way there are 53467 possible subsequences of length 1 (i.e. one for each of the 53467 edges present in WN1.5). The fact that for Spanish the number of partial chains of length 1 is 53467 too, simply means that Spanish wordnet covers all of these 11 tops. In the case of Dutch the figure is a little lower and this means that at least one WN1.5 top is not covered by Dutch wordnet. The corresponding figures for verbs are lower simply because the number of WN1.5 tops is greater (573) and the degree of coverage for the different wordnets is obviously lower.

So, we must find a compromise between the degree of coverage and the significance of such coverage. The first rows present a high coverage but the significance is very low. As the length grows, the significance of the overlapping is greater but the coverage is poor. For nouns the interesting and useful figures can be found in the rows corresponding to lengths 3 to 7. In the case of verbs useful information can be obtained from lengths 2 and 3.

When considering the edge coverage we must taken into account that covering an edge means covering the adjacent nodes and the relation between them. So interesting information could be extracted from rows corresponding to edge lengths 2 or 3, for nouns, and 1 for verbs.

How to use this information? We can, by means of the verbose mode, select all the subchains of length 3 to 7 covered by 2 languages and try to complete these chains for the other language. For instance, for nouns, we can select the 901 chains of length 7 covered over WN1.5 by Spanish and Dutch wordnets. From these, 228 are covered too by Italian, so
only 673 chains must be checked for Italian. Fortunately most of these chains own a common prefix (i.e. some of the initial nodes of the chains are common to several of them). So the amount of work to be done is limited. In this way the information can be used for guiding the construction of subset2 trying to improve the overlapping between chains. Similar considerations could be pointed out for edge chains. In Appendix III, the tables are given for when the wordnets are projected over Dutch, Spanish and Italian. The results and conclusions are similar to the above.

Considering subsequences, as has been pointed out above, is a very useful source of information for 1) assessing the quality of coverage of the wordnets and 2) developing criteria for guiding the extension of the wordnets and obtaining data for supporting such criteria. Another complementary source of information consists of the subsequences containing gaps. We only discuss the results of overlapping subsequences with one or two gaps. \(^{11}\) For one gap, all the tables are presented, for two gaps only the two tables corresponding to partial noun coverage projected over the WN1.5 structure are presented. The reason is simply that the conclusions are basically the same when projecting over the other wordnets and that for verbs the figures when dealing with rather long chains have little significance.

The following tables account for the overlapping of partial chains containing one gap (node vs edge, noun vs verb) projected over WN1.5, Dutch, Italian and Spanish WN structure, for different lengths of the chain.

Table 52: Coverage of partial noun chains of NODES with 1 gap projected over WN1.5 structure

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>NL</th>
<th>∩(ES, NL)</th>
<th>∩(ES, IT)</th>
<th>∩(IT, NL)</th>
<th>∩(ES, IT, NL)</th>
<th>WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>9553</td>
<td>5750</td>
<td>16836</td>
<td>15953</td>
<td>5460</td>
<td>5709</td>
<td>5672</td>
<td>53434</td>
</tr>
<tr>
<td>4</td>
<td>9293</td>
<td>4647</td>
<td>14210</td>
<td>13107</td>
<td>4314</td>
<td>4193</td>
<td>4127</td>
<td>52913</td>
</tr>
<tr>
<td>5</td>
<td>8742</td>
<td>3281</td>
<td>9753</td>
<td>9187</td>
<td>2988</td>
<td>2932</td>
<td>2901</td>
<td>50693</td>
</tr>
<tr>
<td>6</td>
<td>7721</td>
<td>1541</td>
<td>6373</td>
<td>5924</td>
<td>1284</td>
<td>270</td>
<td>227</td>
<td>45029</td>
</tr>
<tr>
<td>7</td>
<td>5831</td>
<td>589</td>
<td>2333</td>
<td>2041</td>
<td>496</td>
<td>12</td>
<td>11</td>
<td>32299</td>
</tr>
<tr>
<td>8</td>
<td>3853</td>
<td>147</td>
<td>664</td>
<td>499</td>
<td>103</td>
<td>3</td>
<td>3</td>
<td>20558</td>
</tr>
<tr>
<td>9</td>
<td>1682</td>
<td>35</td>
<td>190</td>
<td>78</td>
<td>19</td>
<td>3</td>
<td>3</td>
<td>11821</td>
</tr>
<tr>
<td>10</td>
<td>654</td>
<td>12</td>
<td>60</td>
<td>18</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>5881</td>
</tr>
<tr>
<td>11</td>
<td>164</td>
<td>1</td>
<td>16</td>
<td>10</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2576</td>
</tr>
<tr>
<td>12</td>
<td>61</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1176</td>
</tr>
<tr>
<td>13</td>
<td>7</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>659</td>
</tr>
<tr>
<td>14</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>446</td>
</tr>
</tbody>
</table>

Table 53: Coverage of partial noun chains of EDGES with 1 gap projected over WN1.5 structure

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>NL</th>
<th>∩(ES, NL)</th>
<th>∩(ES, IT)</th>
<th>∩(IT, NL)</th>
<th>∩(ES, IT, NL)</th>
<th>WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>1260</td>
<td>2969</td>
<td>2903</td>
<td>1192</td>
<td>334</td>
<td>309</td>
<td>52913</td>
</tr>
<tr>
<td>4</td>
<td>0</td>
<td>281</td>
<td>1478</td>
<td>1369</td>
<td>270</td>
<td>133</td>
<td>123</td>
<td>50693</td>
</tr>
<tr>
<td>5</td>
<td>0</td>
<td>50</td>
<td>273</td>
<td>184</td>
<td>47</td>
<td>45</td>
<td>45</td>
<td>45029</td>
</tr>
<tr>
<td>6</td>
<td>0</td>
<td>4</td>
<td>33</td>
<td>15</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>32299</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>0</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>20558</td>
</tr>
</tbody>
</table>

Table 54: Coverage of partial VERB chains of NODES with 1 gap projected over WN1.5 structure

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>NL</th>
<th>∩(ES, NL)</th>
<th>∩(ES, IT)</th>
<th>∩(IT, NL)</th>
<th>∩(ES, IT, NL)</th>
<th>WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>826</td>
<td>954</td>
<td>822</td>
<td>642</td>
<td>899</td>
<td>347</td>
<td>317</td>
<td>6383</td>
</tr>
<tr>
<td>4</td>
<td>486</td>
<td>111</td>
<td>317</td>
<td>220</td>
<td>104</td>
<td>2</td>
<td>2</td>
<td>3853</td>
</tr>
<tr>
<td>5</td>
<td>329</td>
<td>20</td>
<td>82</td>
<td>50</td>
<td>18</td>
<td>0</td>
<td>0</td>
<td>1894</td>
</tr>
<tr>
<td>6</td>
<td>121</td>
<td>0</td>
<td>8</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>865</td>
</tr>
<tr>
<td>7</td>
<td>39</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>403</td>
</tr>
<tr>
<td>8</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>153</td>
</tr>
</tbody>
</table>

\(^{11}\) It is also possible to consider gaps of 3 or more nodes and even 0-gaps subsequences. The 0-gaps subsequences have gaps at the beginning and/or at the end of the chain. The usefulness of these sequences is however lower.
Table 55: Coverage of partial VERB chains of EDGES with 1 gap projected over WN1.5 structure

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>NL</th>
<th>∩(ES, NL)</th>
<th>∩(ES, IT)</th>
<th>∩(IT, NL)</th>
<th>∩(ES, IT, NL)</th>
<th>WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>0</td>
<td>11</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>3853</td>
</tr>
</tbody>
</table>

Table 56: Coverage of partial noun chains of NODES with 1 gap projected over Dutch wordnet

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>∩(ES, IT)</th>
<th>NL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>3694</td>
<td>3457</td>
<td>3428</td>
<td>12256</td>
</tr>
<tr>
<td>4</td>
<td>3543</td>
<td>1279</td>
<td>1243</td>
<td>11747</td>
</tr>
<tr>
<td>5</td>
<td>2987</td>
<td>471</td>
<td>456</td>
<td>10324</td>
</tr>
<tr>
<td>6</td>
<td>2222</td>
<td>222</td>
<td>221</td>
<td>8018</td>
</tr>
<tr>
<td>7</td>
<td>1333</td>
<td>179</td>
<td>177</td>
<td>5362</td>
</tr>
<tr>
<td>8</td>
<td>569</td>
<td>68</td>
<td>62</td>
<td>2654</td>
</tr>
<tr>
<td>9</td>
<td>151</td>
<td>2</td>
<td>0</td>
<td>1003</td>
</tr>
<tr>
<td>10</td>
<td>58</td>
<td>0</td>
<td>0</td>
<td>295</td>
</tr>
<tr>
<td>11</td>
<td>22</td>
<td>0</td>
<td>0</td>
<td>87</td>
</tr>
<tr>
<td>12</td>
<td>8</td>
<td>0</td>
<td>0</td>
<td>31</td>
</tr>
</tbody>
</table>

Table 57: Coverage of partial noun chains of EDGES with 1 gap projected over Dutch wordnet

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>∩(ES, IT)</th>
<th>NL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>1139</td>
<td>7</td>
<td>6</td>
<td>11747</td>
</tr>
<tr>
<td>4</td>
<td>516</td>
<td>0</td>
<td>0</td>
<td>10324</td>
</tr>
<tr>
<td>5</td>
<td>314</td>
<td>0</td>
<td>0</td>
<td>8018</td>
</tr>
<tr>
<td>6</td>
<td>73</td>
<td>0</td>
<td>0</td>
<td>5362</td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>2654</td>
</tr>
</tbody>
</table>

Table 58: Coverage of partial verb chains of NODES with 1 gap projected over Dutch wordnet

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>∩(ES, IT)</th>
<th>NL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>13769</td>
<td>3385</td>
<td>2708</td>
<td>23682</td>
</tr>
<tr>
<td>4</td>
<td>13559</td>
<td>1671</td>
<td>1284</td>
<td>23422</td>
</tr>
<tr>
<td>5</td>
<td>12715</td>
<td>535</td>
<td>387</td>
<td>22584</td>
</tr>
<tr>
<td>6</td>
<td>10638</td>
<td>24</td>
<td>17</td>
<td>21256</td>
</tr>
<tr>
<td>7</td>
<td>8577</td>
<td>0</td>
<td>0</td>
<td>19279</td>
</tr>
<tr>
<td>8</td>
<td>6688</td>
<td>0</td>
<td>0</td>
<td>16995</td>
</tr>
<tr>
<td>9</td>
<td>4245</td>
<td>0</td>
<td>0</td>
<td>14475</td>
</tr>
<tr>
<td>10</td>
<td>2021</td>
<td>0</td>
<td>0</td>
<td>11913</td>
</tr>
<tr>
<td>11</td>
<td>1011</td>
<td>0</td>
<td>0</td>
<td>9144</td>
</tr>
<tr>
<td>12</td>
<td>520</td>
<td>0</td>
<td>0</td>
<td>6632</td>
</tr>
<tr>
<td>13</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>4480</td>
</tr>
</tbody>
</table>

Table 59: Coverage of partial verb chains of EDGES with 1 gap projected over Dutch wordnet

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>∩(ES, IT)</th>
<th>NL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>12</td>
<td>0</td>
<td>0</td>
<td>23422</td>
</tr>
</tbody>
</table>

Table 60: Coverage of partial noun chains of NODES with 1 gap projected over Italian wordnet

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>NL</th>
<th>∩(ES, NL)</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>148</td>
<td>560</td>
<td>513</td>
<td>4067</td>
</tr>
<tr>
<td>4</td>
<td>139</td>
<td>239</td>
<td>226</td>
<td>2405</td>
</tr>
<tr>
<td>5</td>
<td>104</td>
<td>42</td>
<td>35</td>
<td>1960</td>
</tr>
<tr>
<td>6</td>
<td>15</td>
<td>12</td>
<td>10</td>
<td>1035</td>
</tr>
<tr>
<td>7</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>492</td>
</tr>
</tbody>
</table>

Table 61: Coverage of partial noun chains of EDGES with 1 gap projected over Italian wordnet

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>NL</th>
<th>∩(ES, NL)</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>43</td>
<td>68</td>
<td>2</td>
<td>2405</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>1960</td>
</tr>
</tbody>
</table>
### Table 62: Coverage of partial VERB chains of NODES with 1 gap projected over Italian wordnet

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>NL</th>
<th>∩ (ES, NL)</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>443</td>
<td>211</td>
<td>209</td>
<td>612</td>
</tr>
<tr>
<td>4</td>
<td>379</td>
<td>190</td>
<td>131</td>
<td>522</td>
</tr>
<tr>
<td>5</td>
<td>213</td>
<td>81</td>
<td>38</td>
<td>397</td>
</tr>
<tr>
<td>6</td>
<td>57</td>
<td>10</td>
<td>3</td>
<td>196</td>
</tr>
<tr>
<td>7</td>
<td>4</td>
<td>0</td>
<td>0</td>
<td>54</td>
</tr>
</tbody>
</table>

### Table 63: Coverage of partial VERB chains of EDGES with 1 gap projected over Italian wordnet

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>NL</th>
<th>∩ (ES, NL)</th>
<th>IT</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>522</td>
</tr>
</tbody>
</table>

### Table 64: Coverage of partial noun chains of NODES with 1 gap projected over Spanish wordnet

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>NL</th>
<th>IT</th>
<th>∩ (NL, IT)</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5410</td>
<td>1931</td>
<td>1859</td>
<td>16658</td>
</tr>
<tr>
<td>4</td>
<td>4508</td>
<td>1577</td>
<td>1381</td>
<td>15897</td>
</tr>
<tr>
<td>5</td>
<td>3028</td>
<td>1026</td>
<td>986</td>
<td>14219</td>
</tr>
<tr>
<td>6</td>
<td>1698</td>
<td>361</td>
<td>67</td>
<td>11131</td>
</tr>
<tr>
<td>7</td>
<td>699</td>
<td>180</td>
<td>8</td>
<td>7121</td>
</tr>
<tr>
<td>8</td>
<td>186</td>
<td>42</td>
<td>1</td>
<td>3556</td>
</tr>
<tr>
<td>9</td>
<td>49</td>
<td>9</td>
<td>1</td>
<td>1352</td>
</tr>
<tr>
<td>10</td>
<td>9</td>
<td>1</td>
<td>0</td>
<td>503</td>
</tr>
<tr>
<td>11</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>159</td>
</tr>
</tbody>
</table>

### Table 65: Coverage of partial noun chains of EDGES with 1 gap projected over Spanish wordnet

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>NL</th>
<th>IT</th>
<th>∩ (NL, IT)</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>997</td>
<td>341</td>
<td>122</td>
<td>15897</td>
</tr>
<tr>
<td>4</td>
<td>466</td>
<td>91</td>
<td>34</td>
<td>14219</td>
</tr>
<tr>
<td>5</td>
<td>76</td>
<td>14</td>
<td>9</td>
<td>11131</td>
</tr>
<tr>
<td>6</td>
<td>10</td>
<td>5</td>
<td>0</td>
<td>7121</td>
</tr>
</tbody>
</table>

### Table 66: Coverage of partial VERB chains of NODES with 1 gap projected over Spanish wordnet

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>NL</th>
<th>IT</th>
<th>∩ (NL, IT)</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>205</td>
<td>263</td>
<td>101</td>
<td>1432</td>
</tr>
<tr>
<td>4</td>
<td>87</td>
<td>41</td>
<td>2</td>
<td>756</td>
</tr>
<tr>
<td>5</td>
<td>21</td>
<td>4</td>
<td>0</td>
<td>320</td>
</tr>
<tr>
<td>6</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>100</td>
</tr>
</tbody>
</table>

### Table 67: Coverage of partial VERB chains of EDGES with 1 gap projected over Spanish wordnet

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>NL</th>
<th>IT</th>
<th>∩ (NL, IT)</th>
<th>ES</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>5</td>
<td>0</td>
<td>0</td>
<td>756</td>
</tr>
</tbody>
</table>
The following tables account for the overlapping of partial noun chains containing two gaps (node vs edge) projected over WN1.5. Tables with more gaps are not included here but can be easily generated, in the way described in 3.2.2.

**Table 68: Coverage of partial noun chains of NODES with 2 gaps projected over WN1.5 structure**

<table>
<thead>
<tr>
<th>LENGTH</th>
<th>ES</th>
<th>IT</th>
<th>NL</th>
<th>∩(ES, NL)</th>
<th>∩(ES, IT)</th>
<th>∩(IT, NL)</th>
<th>∩(ES, IT, NL)</th>
<th>WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>4</td>
<td>7206</td>
<td>1654</td>
<td>6337</td>
<td>5620</td>
<td>1546</td>
<td>782</td>
<td>662</td>
<td>52913</td>
</tr>
<tr>
<td>5</td>
<td>2713</td>
<td>1548</td>
<td>6729</td>
<td>5939</td>
<td>1423</td>
<td>899</td>
<td>793</td>
<td>50693</td>
</tr>
<tr>
<td>6</td>
<td>2521</td>
<td>916</td>
<td>5502</td>
<td>4731</td>
<td>806</td>
<td>478</td>
<td>396</td>
<td>45029</td>
</tr>
<tr>
<td>7</td>
<td>2172</td>
<td>495</td>
<td>2176</td>
<td>1676</td>
<td>395</td>
<td>129</td>
<td>119</td>
<td>32299</td>
</tr>
<tr>
<td>8</td>
<td>1662</td>
<td>323</td>
<td>1249</td>
<td>1099</td>
<td>273</td>
<td>36</td>
<td>28</td>
<td>20558</td>
</tr>
<tr>
<td>9</td>
<td>1239</td>
<td>136</td>
<td>191</td>
<td>157</td>
<td>100</td>
<td>0</td>
<td>0</td>
<td>11821</td>
</tr>
<tr>
<td>10</td>
<td>606</td>
<td>30</td>
<td>50</td>
<td>50</td>
<td>26</td>
<td>0</td>
<td>0</td>
<td>5881</td>
</tr>
<tr>
<td>11</td>
<td>271</td>
<td>2</td>
<td>9</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>2576</td>
</tr>
<tr>
<td>12</td>
<td>179</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1176</td>
</tr>
<tr>
<td>13</td>
<td>115</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>659</td>
</tr>
<tr>
<td>14</td>
<td>25</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>446</td>
</tr>
<tr>
<td>15</td>
<td>2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>82</td>
</tr>
</tbody>
</table>

The same considerations pointed out when presenting the partial chains are valid here. Obviously the chains are longer in this case (e.g. for having a subchain with one node gap the minimum length is 3, with 2 gaps the minimum length is 4).

A possible useful criterion for using this information could be simply generating the set of sequences belonging to the intersection of all languages and having a gap and then from this set generating for each language $L$ the subset of sequences where this gap is not covered for $L$. For instance, for nouns, the row corresponding to length 5 chains shows that the intersection consists of 2901 sequences with 1 gap. In the worst case all the gaps correspond to every language, i.e. the gaps are not covered by any of the 3 languages other than English, and all the gaps correspond to different ILI nodes. In this worst case the task would consist of filling these 2901 gaps for each language. Doing so, the number of length 5 noun chains without gaps will increase from 780 to 780 + 2901 = 3681 with an increment of 470%! If we consider also the length-5 noun chains with 2 gaps, another increment (in this case of 793 chains) can be obtained.

A careful evaluation of the most promising chains in order to optimize the human resources must be done, but we are sure that using the information provided by these tables and generating the corresponding occurrences (with the verbose option) can be helpful as a guideline for extending and improving the wordnets.
4. Updating the ILI

The ILI will be updated in 3 ways (see [Peters et al, fc] for more details):

1. Improving the information given for the current ILI-records: e.g. adding missing glosses
2. Adding missing concepts or gaps in the ILI occurring in the other wordnets
3. Adding globalized sense, grouping closely related senses of words in the ILI

Gaps (2) will be added later in the project. For Subset1 we have focussed on improving the glosses (1) and creating globalized sense-groups (3). The latter are needed to deal with the reduction of the high level of granularity of the WordNet sense distinctions and the inconsistent treatment of regular polysemy in lexical resources. As explained in [Peters et al, 1998], differences in the sense-distinctions across resources may lead to mismatches or fuzzy-matching across wordnets. For example, university may be used to refer to both the institute and the building but we see that resources often only represent one of these meanings, or conflate them in a single meaning. This may again result in a situation that the local synsets for university cannot be matched across wordnets.

To limit this danger, we extend the ILI with globalized senses that represent sets of more specific but related senses of the same word. Three main types of relations are distinguished:

1. metonymy, e.g. grouping building-institute senses
2. generalization, e.g. grouping specific uses of a single more general sense
3. diathesis alternation, e.g. grouping causative/inchoative senses

In Figure 5, we see that the original linking of Dutch, Italian and Spanish equivalents for university has been extended with an HAS_EQ_METONYM relation to a new globalized ILI-record university which contains a reference to two more specific meanings. Via the HAS_EQ_METONYM relations the synsets can be retrieved despite of the different ways in which they are linked to the more specific synsets. It is not necessary that the metonymy-relation also holds in the local language. In this example only the Dutch wordnet has two senses that parallel the metonymy-relation in the ILI.12 The Italian and Spanish example only list one sense (which may be correct or an omission in their resources). In the case of Spanish there are multiple equivalences to both senses of university, whereas the Italian synset is only linked to the building sense. The Spanish example is in fact equivalent to the new globalized ILI-record.

Similar globalized records are added for generalizations or verbal alternations such as causative and non-causative meanings: he opens the door, versus the door opens [Levin 1993]. In that case an HAS_EQ_GENERALIZATION or HAS_EQ_DIATHESIS relation will hold for synsets linked to more specific ILI-records that can be grouped in these ways. The generation of these equivalence relations is done fully automatically. After extending the ILI with more global concepts, the HAS_EQ_METONYM, HAS_EQ_GENERALIZATION or HAS_EQ_DIATHESIS will be automatically generated for all synsets which have at least one of the specific ILI-records in the globalized ILI-records as the target of an EQ_SYNONYM or EQ_NEAR_SYNONYM relation. There is no need for the local wordnet builders to consider each of these equivalence-extensions manually.

---

12 The relation between these two Dutch senses is now also expressed via the metonymy-equivalence relation to the more global ILI-record. The globalized ILI-record may also create metonymic relations between different forms which represent the same semantic relation, such as universiteit (university institute) and universiteitsgebouw (university building) in Dutch.
4.1 Clustering Methods

This section describes the clustering methods which have been and still are being applied in order to identify these sense groups.

4.1.1 Manual clustering

We started off with a manual examination of the polysemous words in the Base concept (BC) set and their senses which had originally been rejected in the BC selection process, but had been selected by at least two individual language partners. This set will be referred to below as the Rejected Concepts (RC). These originally rejected synsets were evaluated against the base concept set, and their possible inclusion into the BC set was investigated. Three different strategies has been applied in order to select relevant members from the RC set as new BCs.

1) The first strategy uses the average number of semantic relations selected noun base concepts have in WordNet1.5 (19.49) as the selection threshold for rejects. The following selection criteria have been applied:
   - Because of the high number of direct hyponymic relations within the BC (636) direct RC hyponyms of existing BCs have principally not been selected. For instance, ‘airplane’ has not been selected because of its NBC direct hypernym ‘aircraft’. ‘rate’ (a magnitude relative to a time unit; “they traveled at a rate of 55 miles per hour” or “the rate of change was faster than expected”) is subsumed by BC ‘relation’ (an abstraction belonging to or characteristic of two entities or parts together) and has not been selected.
   - Basic level concepts (43) like bed, wheel, shoe, window, glass, eye, soup, pants, antilope represent a level of lexicalisation which is too specific, and have not been selected.
   - Taxonomic terms within the field of biology have not been selected as new BCs. They have very specific technical meanings, and are subsumed by the BC ‘group’.

This first selection task yielded 14 new potential BC members.

2) The second selection method consisted of the manual examination of rejected synsets which share a word form member with one or more NBC synsets. Only word forms shared by at least four synsets have been taken into account and it has been investigated whether the rejected synsets belonging to these sense groups did not all originate from only one language specific wordnet, but have a more or less even distribution over the different language sites. This method resulted in 100 new BC members.
3) The third strategy used the metric developed by [Agirre & Rigau, 1996] for computing conceptual distance between WordNet nodes representing different senses of the same word. Using a threshold of .3 yielded 21 RC synsets as candidates for BC membership.

For the manual identification of encoding of metonymic regularities between senses the following aspects of systematic polysemy apply amongst others [Apresjan, 1974], [Pustejovsky 1995]:

- a general notion of involvedness: the senses are related within a typical situation; e.g. social group versus belief, organisation vs building; result, e.g activity vs product;
- constituent or portion/part vs, whole relations, e.g wood-tree, person-social group
- function e.g. liquid-beverage

For the identification of generalization it was difficult to find clear cut criteria. The general criteria used included the level of fine-grainedness of the sense distinctions and the possibility to make an ontological generalization over the senses involved, which constitutes a lowest common denominator that all grouped senses share.

This first manual clustering round resulted in 31 verb and 119 noun groups.

4.1.2 (Semi-)automatic clustering

In addition to the manual clustering, various automatic clustering methods have been examined. Most of these methods rely on the internal hierarchical organization of WordNet and, except for autohyponymy (see section 4.1.2.2), they are all used in the WordNet interface to compute semantic similarity. With respect to using external resources to aid clustering, we have only looked at CoreLex (section 4.1.2.5). However, we envisage using other existing lexical resources, such as machine-readable dictionaries and ontological classifications. Thus far, we have limited ourselves to homographs of the same part of speech. The methods are described in the following paragraphs.

4.1.2.1 Sisters

Word senses that share the same hypernym are called sisters. In the example below, both senses of table have furniture as their direct hypernym:

- table-2
  'a piece of furniture having a smooth flat top supported by one or more vertical legs; "it was a sturdy table"
  
- table-3
  'a piece of furniture with tableware for a meal laid out on it; "I reserved a table at my favorite restaurant"

Using the sister criterion generates patterns of generalization. In the example above, the given senses can be used to refer to the same object, highlighting different aspects of it. However, in some cases the clustered senses refer to different objects in the real world. This is illustrated by the following example, where all three senses share the direct hypernym vine.

- butterfly pea
  - 'vine of tropical Asia having pinnate leaves and bright blue yellow-centered flowers'
  - 'large-flowered wild twining vine of SE and C US having pale blue flowers'
  - 'large-flowered weakly twining or prostrate vine of NJ to tropical E N America, sometimes cultivated for its purple and white flowers'

As these senses denote different species, they are not near-synonyms. However, they are very similar in nature, and can be clustered on that basis. It must be taken into account that, and this is true for all generalizations, the meanings cannot be used interchangeably. The most specific semantic content these particular senses share is the meaning of the direct hypernym.

---

13 SHE has also provided manually identified metonymy and generalization relations for the semantic fields Building and Institute, which resulted in 33 new groups. This was done to investigate effectiveness of metonymic clusters.

14 The sister relation is not limited to two senses, but can also occur between three or more senses of the same word. Sometimes, a particular word exhibits more than one type of sister relation.
4.1.2.2 Autohyponymy
The term *autohyponymy* is used to refer to words whose senses are each others direct hypernyms or hyponyms (Cruse, 1986). Sharing the same hypernymic chain (except for the first node) provides us with a number of combinations where the meanings are very similar and clustering results in homogenous groups. Look at the following examples, where the first sense is the most specific one:

- **variety-3, species**
  'a specific kind of something: "a species of molecule" or "a species of villainy"

- **variety-6, kind, sort, form**
  'a category of things distinguished by some common characteristic or quality; "sculpture is a form of art"; "what kinds of desserts are there?"

- **understand-3, read, interpret, translate**
  'make sense of a language; "She understands French"; "Can you read Greek?"

- **understand-1**
  'know and comprehend the nature or meaning of; "She did not understand her husband"; "I understand what she means"

As this method also leads to generalization clusters it is the meaning of the hypernym synset that can be used to characterize the resulting sense cluster. The specific sense is subsumed by the general one; the hyponym carries extra meaning which is not shared by its parent and/or is typically used in a specific domain.

4.1.2.3 Twins
Twins are synsets that have at least three members in common as the example below illustrates. Their meanings are defined by ‘of rules or patterns’ and ‘act in disregard of laws and rules’, respectively.

- **violate, fail to agree with, go against, break-13, be in violation of**
- **violate, go against, breach, break-6, be in violation of**

This example seems to validate clustering on the basis of the twin criterion. However, some of the twin groupings are more problematic. The synsets below have the following incompatible glosses: ‘motion that does not entail a change of location; “the reflex movements of his eyebrows revealed his surprise”’ and ‘the act of changing your location from one place to another’.

- **change of position, motion, movement, move-3**
- **change of location, motion, movement, move-4**

A number of synsets are linked by a twin relation only because they contain spelling variants, such as *sestet, sextet, sextette*. As we have not yet examined the twin relation in great detail, we cannot fully assess the validity of this method. However, it seems that even in cases where synsets only share two members, this can also be an indication that clustering is possible. An example is travel-4, journey and travel-2, journey, where the meanings are very closely related.

4.1.2.4 Cousins
WordNet1.5 contains a list of 105 node top pairs whose hyponyms exhibit a specific relation to each other (see WordNet database documentation on groups, file groups.715). These pairs have been identified and listed by lexicographers. The treatment of these so-called cousins is still in its experimental stage; the resulting list is incomplete and does not offer a consistent and structured list of recurrent patterns between sets of words. Examples of cousin relations are container-containerful and food-tableware, listed below.

- **container-1**
  'something that holds things, especially for transport or storage'

- **containerful-1**
  'the quantity that a container will hold'

---

15 This documentation is included in the WordNet database, downloadable from http://www.princeton.edu/~wn/
food-1, nutrient
‘any substance that can be metabolized by an organism to give energy and build tissue’
tableware-1
‘articles for use at the table’

Looking at the first pair, there are a large number of words that occur both as hyponyms of the container node and the containerful node, such as bag, can, cup, glass, shovel, spoon and thimble. These are all good examples of the regular polysemic pattern that exists between container and containerful. On further investigation, we find that the cousin relation is not limited to senses sharing a word form. For example, WordNet contains no words that have both a food and a tableware meaning. While words such as silver plate, gold plate, crockery and chop sticks all occur as hyponyms of tableware, they are not found in a food sense. Cousin relations, thus, do not necessarily generate regular polysemous patterns, but sometimes capture semantic relations between words of a more schema-like nature. Within the scope of the present research we are only interested in sense distinctions of individual words and can only use those cousin relations generating clusters that share word forms.

4.1.2.5 CoreLex
An attempt at making systematic polysemic patterns in WordNet explicit has been made by Buitelaar (1998). The CoreLex database\(^{16}\) contains 126 semantic types, covering 39,937 nouns in 317 systematic polysemous classes. Three steps were taken to derive CoreLex from WordNet. Firstly, all polysemous nouns in WordNet were reduced to a set of Basic Types, corresponding largely to WordNet’s ‘unique beginners’ and ‘top nodes’, such as artifact, causal agent, shape and act. Subsequently, systematic groupings of nouns were created on the basis of their Basic Types distributions. For example, the noun banana, occurring both in a food and a plant sense, was put in a group with other nouns exhibiting the same pattern, such as coriander, grapefruit, plantain and mulberry. The final step consists of integration into the Core Lexical Engine [Pustejovsky, 1995].

On examining the polysemous classes, we found a number of disadvantages to the CoreLex system. Firstly, 19 of them consist of only one Basic Type and therefore do not display systematic polysemy. More importantly, the generated classes are not always homogeneous in nature; particularly the larger groups do not necessarily exhibit regular polysemic patterns and occurrences of ‘monsters’ are not infrequent. Often there is scope for further subclustering. For example, we find bundle, package, packet, ragbag, deck, edition, library, menagerie, repertory belonging to the same CoreLex type (arg, a combination of the Basic Types artifact and group) where we find the first 4 words covered by the more specific hyponymic nodes collection-1 and container-1 and the last three by collection-1 and facility-1. For our purposes, the main problems with CoreLex are caused by the fact that the Basic Types are largely based on very high-level nodes in the WordNet hierarchy. In order to obtain more homogeneous classes, we propose to examine recurrent distributional patterns at a more specific level in the hierarchy.

4.2 Testing automatically created sense groups
We have performed a first validation test of automatic procedures for deriving sense clusters. For this purpose we carried out an experiment in which different fragments of the Dutch and Spanish wordnets were compared, both before and after extending the ILI with composite ILI records. For the experiment, composite ILLs have been generated automatically on the basis of two methods:

- We selected a number of metonymic relations and subsequently extracted all words that have one sense occurring as a (sub)hyponym of one element of the relation and another sense as a (sub)hyponym of the other element. Some of these relations feature in the cousin table, discussed in section 4.2.4. As suggested in section 4.2.5, the selected relations generally consist of hyponymic nodes that are more specific than WordNet’s top nodes and unique beginners. This method generates regular polysemic patterns.

- From the words selected by the above-mentioned method, we clustered those word senses that are (sub)hyponyms of one of the members of the metonymic relations selected in this experiment. This method extracts generalization clusters and extends the sister relation discussed in 4.2.1 to include those senses that are not direct hyponyms of the shared hyponymic node, i.e. senses that are not co-hyponyms. This method also subsumes autohyponomy.

---
\(^{16}\) Available from http://www.cs.brandeis.edu/~paulb/CoreLex/overview.html
Table 70 gives the totals for the extracted records. In total 700 new composite ILI-records have been added (214 metonymic groupings and 486 generalization pairs), involving 1557 ILI-records. Note that this method is fully automatic and can easily be extended to all senses in WordNet1.5. In general, we see here that the largest metonymic classes are *animal/food* and *plant/food*. The largest set of generalization is extracted for *move*. After extending the ILI with the new concepts, the equivalence relations of the Spanish and Dutch wordnet to the ILI have been updated. This is done automatically by the database: any synset that is related to an ILI-record included in a composite ILI will get an additional metonymy or generalization link to this composite ILI-record. For the Dutch wordnet 602 links have been added, and for the Spanish wordnet 521 links.

Table 70: Automatic derived generalizations and metonymy-relations

<table>
<thead>
<tr>
<th>Semantic Class</th>
<th>Total Descendants</th>
<th>Generalization Clusters</th>
<th>Metonymic Clusters</th>
<th>Intersecting Metonymy Class</th>
<th>Total Composite s</th>
<th>Percentual Coverage of all Senses</th>
</tr>
</thead>
<tbody>
<tr>
<td>animal</td>
<td>3842</td>
<td>80</td>
<td>81</td>
<td>food</td>
<td>161</td>
<td>4.19%</td>
</tr>
<tr>
<td>food</td>
<td>4750</td>
<td>48</td>
<td>100</td>
<td>food</td>
<td>148</td>
<td>3.11%</td>
</tr>
<tr>
<td>plant</td>
<td>2123</td>
<td>64</td>
<td>181</td>
<td>animal/plant</td>
<td>245</td>
<td>11.54%</td>
</tr>
<tr>
<td>organization</td>
<td>846</td>
<td>31</td>
<td>25</td>
<td>construction</td>
<td>56</td>
<td>6.61%</td>
</tr>
<tr>
<td>construction</td>
<td>1210</td>
<td>81</td>
<td>25</td>
<td>organization</td>
<td>106</td>
<td>8.76%</td>
</tr>
<tr>
<td>move</td>
<td>708</td>
<td>176</td>
<td>8</td>
<td>sound</td>
<td>184</td>
<td>25.98%</td>
</tr>
<tr>
<td>sound</td>
<td>192</td>
<td>6</td>
<td>8</td>
<td>move</td>
<td>14</td>
<td>7.29%</td>
</tr>
<tr>
<td>Total</td>
<td>13671</td>
<td>486</td>
<td>214</td>
<td></td>
<td>914</td>
<td>6.68%</td>
</tr>
</tbody>
</table>

To measure the effect, we mapped Spanish (ES) and Dutch (NL) fragments before and after extending the ILI with these records. All descendants of Dutch and Spanish representatives of the above classes were selected, e.g. all (sub)hyponyms of *bouwwerk-1* (construction) in Dutch and *construcción-4* (construction) in Spanish. In the EuroWordNet database, it is possible to 'project' these language-specific descendant word meanings to the other language (translate via the ILI). The result is a set of word meanings in the target language connected to the source language meanings via ILI-records. By taking the intersection of this projection in both directions we get an idea of the overlap of these semantic clusters (for further details, see (Peters et al., forthcoming)).

Table 70 gives the results of this mapping in both directions for each hierarchical node, once before the ILI-extension (rows headed by ILI-0) and once after the update (rows headed by ILI-1). For each language, the first column gives the total number of (sub)hyponyms per hierarchical node (the descendants), the second column gives the number of word meanings that have been projected to that particular language (from Spanish to Dutch and from Dutch to Spanish) and the third column lists the percentages of the projection for the total set of descendant word meanings. The last two columns give, for each language, the intersection of the projected word meanings (WMs in table above and the descendant word meanings, in absolute numbers and percentages. The bottom rows list the totals.

The general tendency for the Dutch wordnet is that the projection increased by 5.8%, whereas the increase of the intersection is 2.25%. For the Spanish wordnet these figures are 3.13% and 2.01% respectively. If we compare the increase of the projection (103 word meanings for Dutch and 56 for Spanish) with the increase in intersection (40 word meanings for Dutch and 36 for Spanish), we see that between 40-65% of the extended projection is effective, i.e. leads to an increase of the intersection. We suspect that the remaining incompatibilities either reflect a real difference in coverage or are caused by diverging classifications (e.g. *milk* is classified as a *product* instead of *comestible*; a hypernym of *food*).

---

17 In the case of animal and food, we have concentrated on the metonymic patterns. Because of the size of both sets, we have not investigated the instances of generalization.

18 In some cases, the projection extends the total set (more than 100%). This means that these words have been classified differently in the target language of the projection.
Table 71: Projection and Intersection increase Dutch-Spanish after adding sense-clusters to the ILI

<table>
<thead>
<tr>
<th></th>
<th>Projection to the Dutch WordNet</th>
<th>Projection to the Spanish WordNet</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Desc. WM</td>
<td>Projection</td>
</tr>
<tr>
<td></td>
<td>WM % of ES</td>
<td>WM % of NL</td>
</tr>
<tr>
<td>organization</td>
<td>ILL-0</td>
<td>48</td>
</tr>
<tr>
<td></td>
<td>ILL-1</td>
<td>48</td>
</tr>
<tr>
<td>construction</td>
<td>ILL-0</td>
<td>344</td>
</tr>
<tr>
<td></td>
<td>ILL-1</td>
<td>344</td>
</tr>
<tr>
<td>food</td>
<td>ILL-0</td>
<td>154</td>
</tr>
<tr>
<td></td>
<td>ILL-1</td>
<td>154</td>
</tr>
<tr>
<td>move</td>
<td>ILL-0</td>
<td>1183</td>
</tr>
<tr>
<td></td>
<td>ILL-1</td>
<td>1183</td>
</tr>
<tr>
<td>sound</td>
<td>ILL-0</td>
<td>47</td>
</tr>
<tr>
<td></td>
<td>ILL-1</td>
<td>47</td>
</tr>
<tr>
<td>Total</td>
<td>ILL-0</td>
<td>1776</td>
</tr>
<tr>
<td></td>
<td>ILL-1</td>
<td>1776</td>
</tr>
<tr>
<td>Increase</td>
<td></td>
<td>103</td>
</tr>
</tbody>
</table>

If we examine the figure in more detail, we see the following tendencies:

- the methodology is effective for organization, construction and move;
- the methodology is hardly effective for food and sound;

In the case of move (see table 70) we can expect that the effect is high because the extension (the composite ILIs) already makes up 25% of the total set of descendant senses. In the case of organization and construction, it is more remarkable because the extension only makes up 6-8% of the total of descendants. Further inspection shows that the effect for construction and organization is evenly spread over metonymy and generalization (50-70%) whereas the effect for move is almost exclusively due to generalization (97%). The fact that the effect is small for sound is in line with the low extension with composite ILI-records (6% of the total number of descendants). For food, the effect is more disappointing, given the much higher proportion of composite ILIs (11%).

To verify the quality of the extension, we have manually inspected the new word meanings that were projected from the Spanish wordnet to the Dutch wordnet. This inspection showed hardly any projections that are incompatible with the classifications of the projection before the extension, except for those that fall within the metonymic extension. In so far as there is a degree of variation in classification across the wordnets ([Peters et al., forthcoming], the extension is not worsening this effect. However, there is metonymic over-generation across the wordnets:

Table 72: Errors generated by automatically derived Composite ILIs

<table>
<thead>
<tr>
<th></th>
<th>New Projections to NL after the ILI Extension</th>
<th>Metonymic Overgeneration</th>
<th>Genuine Errors</th>
</tr>
</thead>
<tbody>
<tr>
<td>food</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>construction</td>
<td>16</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>organization</td>
<td>19</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>move</td>
<td>65</td>
<td>0</td>
<td>4</td>
</tr>
</tbody>
</table>

Metonymic over-generation was to be expected, since regular polysemy does not necessarily hold across the languages. It may be caused by a cultural difference (e.g. not all plants and animals are considered to be food in all language/cultures), although we did not find any examples of this type of over-generation. Another possible reason for over-generation is a difference in lexicalization (e.g. metonymic meanings can be lexicalized by different word forms). In the case of plant/food, there is only one occurrence of over-generation: in the compound vanilleplant (the plant from which vanilla is extracted) the headword plant blocks the spice interpretation. The same phenomenon occurs more often with organization/construction, because a number of Dutch compounds can only refer to a building, such as vestigingswerk (defense construction), and verenigingsgebouw (building where the club is seated). Among the constructions we find several genuine cases of over-generation: gemeenschap (the community), godsdienst (religion), delegatie (delegate), commissie (commission) are all groups of people without an associated building. Finally, in the
case of move, three errors occur: 

bidden (to pray), gelijkspeLEN (to finish a game with even scores) and verschrijven (to make a mistake with writing). However, these are due to incorrect translations or dubious classifications which also occur within non-extended projections. Metonymic over-generation is not problematic since it is up to the builder of the local wordnet to decide whether to include the metonymic pattern for a particular language. For example, in the cases discussed above the Dutch wordnet will only have an eq_synonym relation with one of the senses related by metonymy, while in other languages we may find the same word linked to multiple meanings.

5 Conclusions
In this document, we have described the work carried out for the first subset in EuroWordNet. The subset has been constructed starting from the common set of Base Concepts, including all the major synsets on which the other concepts depend. This set has been extended top-down by each site separately. The resulting wordnets have been described in terms of quantity (entries, senses, synsets, language-internal relations and equivalence relations), by comparing the overlap with Parole lexicons and by measuring the conceptual coverage by clustering of Top-Ontology concepts. Whereas the Spanish wordnet already has reached full coverage (advancing the planning), the Dutch wordnet has just covered the first subset with a higher density of language internal relations, and the Italian wordnet has full coverage but lacks equivalence relations. The distribution of the wordnets over the top-ontology was surprisingly balanced. Some slight imbalances for 1stOrder Entities have to be corrected. Similarly, the overlap with the top-frequent Parole entries is also very high. Missing entries can easily be added.

In addition we have carried out two comparisons to get an impression of the consistency of the wordnets. Both comparisons showed promising results. The in-depth comparison of 18 fields showed reasonable intersections. Most of the mistakes are due to translation errors. Alternative classifications can be used to encode multiple hyperonym. A similar conclusion has been made from the overall comparison. There is a high degree of overlap between subsequences and sequences with 1 gap. By filling these gaps we can improve the coverage in a coordinated way. Furthermore, extremely tangled graphs (Dutch verbs) are mostly due to generation of wrong translations.

The following improvements will therefore be made to the wordnets in the next building phase:

- improve balancing of 1stOrderClusters (Dutch and Italian)
- extend with missing top-frequent Parole entries (Dutch, Italian, Spanish)
- extend coverage (Dutch)
- check translations of extremely long hyponymy chains (especially Dutch verbs)
- fill sequences with 1 gap (Italian, Spanish and Dutch)
- extend translations (Italian)
- improve translation heuristics (Spanish and Dutch)

Finally, this deliverable describes the work done for updating the Inter-Lingual-Index (ILI) that interconnects the different wordnets. We showed that using fully automatic techniques we can achieve up to 5% improvement in matching across wordnets.
References


Peters, I. and Peters, W., *Extracting Regular Polysemic Patterns in WordNet*. Technical Report, University of Sheffield, United Kingdom


Appendix I In-depth comparison of semantic clusters by different sites

Appendix Ia Comparing to the Dutch wordnet

General Comments

- **Mistakes:** most mistakes in the Dutch wordnet are due to wrong translations. It turns out that taking the best 3 translations generated by heuristics generates too many wrong translations. This will be adjusted to the best 2 translations. Only a few mistakes are due to wrong classifications.
- **Alternative classification:** in many cases parts (e.g. parts of buildings) are classified as subtypes of the wholes in the Reference wordnets: e.g. a *room* is both a *type* of *construction* and a *part* of *construction*. This is a systematic difference with the Dutch wordnet where such parts are systematically classified as a type of part and related to the whole by a meronym relation.
- **Coverage:** the coverage of the Dutch wordnet is less then the other wordnets. This is because the other wordnets have included larger proportions in Subset1.
- **Equivalence matching:** the Spanish wordnet has a direct matching of synsets with the ILI only using eq_syninym relations; the Dutch and Italian wordnet also have other types of equivalence relations.
First Order Entities

**Building**

Major Nodes, hyponyms and equivalence relations

<table>
<thead>
<tr>
<th></th>
<th>Hyperonymic Synsets</th>
<th>No of Hypos All Levels</th>
<th>No of ILIs</th>
<th>EQ_S</th>
<th>EQ_N</th>
<th>EQ_Hyper</th>
<th>EQ_Hypo</th>
<th>EQ_Re</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>{construcción-4}</td>
<td>548</td>
<td>548</td>
<td>548</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>{construzione-1}</td>
<td>194</td>
<td>7</td>
<td>5</td>
<td>2</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>{bouwwerk-1}</td>
<td>344</td>
<td>223</td>
<td>39</td>
<td>188</td>
<td>15</td>
<td>0</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>{construction 4}</td>
<td>1220</td>
<td>1220</td>
<td>1220</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Projection of Reference wordnets to Dutch Source wordnet

<table>
<thead>
<tr>
<th></th>
<th>Reference Hyponyms</th>
<th>No Match in Source WordNet</th>
<th>Matching WMs in Source</th>
<th>Source WMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>548</td>
<td>367</td>
<td>181</td>
<td>276</td>
</tr>
<tr>
<td>IT</td>
<td>194</td>
<td>1</td>
<td>6</td>
<td>21</td>
</tr>
<tr>
<td>NL</td>
<td>344</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>1210</td>
<td>920</td>
<td>290</td>
<td>364</td>
</tr>
</tbody>
</table>

Comparing projections for the Dutch wordnet

<table>
<thead>
<tr>
<th></th>
<th>Projected Source WMs</th>
<th>Union</th>
<th>Intersection</th>
<th>Unique in Reference WN</th>
<th>Unique in Source WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>276</td>
<td>487</td>
<td>133</td>
<td>143</td>
<td>211</td>
</tr>
<tr>
<td>IT</td>
<td>21</td>
<td>350</td>
<td>15</td>
<td>6</td>
<td>329</td>
</tr>
<tr>
<td>NL</td>
<td>344</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>WN15</td>
<td>264</td>
<td>538</td>
<td>170</td>
<td>194</td>
<td>174</td>
</tr>
</tbody>
</table>

Errors in Dutch Source:
- wrong translations: 6
- wrong classifications: 8

Errors in Reference: 0

Alternative classifications:
- movable constructions
- parts of buildings
- institutions

Variant Projection of Unmatched ILI-records from Reference wordnets to Dutch Source Wordnet

<table>
<thead>
<tr>
<th></th>
<th>Unmatched ILIs</th>
<th>All Senses of ILIs</th>
<th>No Match in Source</th>
<th>Matching ILI Source in Source WM</th>
<th>Intersect with Source Projection</th>
</tr>
</thead>
</table>
**Comestibles**

**Major Nodes, hyponyms and equivalence relations**

<table>
<thead>
<tr>
<th></th>
<th>Hyperonymic Synsets</th>
<th>No of Hypos All Levels</th>
<th>No of ILIs</th>
<th>EQ_S</th>
<th>EQ_N</th>
<th>EQ_Hyper</th>
<th>EQ_Hypo</th>
<th>EQ_Rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>{alimento-1}</td>
<td>533</td>
<td>533</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>{cibo-1}</td>
<td>157</td>
<td>51</td>
<td>40</td>
<td>10</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>{voedsel-1}</td>
<td>151</td>
<td>154</td>
<td>27</td>
<td>129</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>{food-1}</td>
<td>2123</td>
<td>2123</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Projection of Reference wordnets to Dutch Source wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Reference Hyponyms</th>
<th>No Match in Source WordNet</th>
<th>Matching WMs in Source</th>
<th>Source WMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>533</td>
<td>410</td>
<td>123</td>
<td>135</td>
</tr>
<tr>
<td>IT</td>
<td>51</td>
<td>30</td>
<td>21</td>
<td>35</td>
</tr>
<tr>
<td>WN15</td>
<td>2123</td>
<td>1923</td>
<td>200</td>
<td>191</td>
</tr>
</tbody>
</table>

**Comparing projections for the Dutch wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Projected Source WMs</th>
<th>Union</th>
<th>Intersection</th>
<th>Unique in Reference WN</th>
<th>Unique in Source WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>135</td>
<td>216</td>
<td>70</td>
<td>65</td>
<td>81</td>
</tr>
<tr>
<td>IT</td>
<td>35</td>
<td>162</td>
<td>24</td>
<td>11</td>
<td>127</td>
</tr>
<tr>
<td>WN15</td>
<td>191</td>
<td>239</td>
<td>103</td>
<td>88</td>
<td>48</td>
</tr>
</tbody>
</table>

**Errors in Dutch Source:**
- wrong translations: 6
- wrong classifications: 0

**Errors in Reference:** 0

**Alternative classifications:**
- natural products such as fruits, grain, corn, seeds
- drinks
- parts

**Variant Projection of Unmatched ILI-records from Reference wordnets to Dutch Source Wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Unmatched ILIs</th>
<th>All Senses of ILIs</th>
<th>No Match in Source</th>
<th>Matching ILI in Source</th>
<th>ILIs of Matching Source WM</th>
<th>Intersection with Source Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>3199</td>
<td>3011</td>
<td>188</td>
<td>248</td>
<td>19</td>
<td></td>
</tr>
</tbody>
</table>
### Container

**Major Nodes, hyponyms and equivalence relations**

<table>
<thead>
<tr>
<th></th>
<th>Hyperonymic Synsets</th>
<th>No Hypos of ILIs</th>
<th>No of ILIs</th>
<th>EQ_Hyperonymy</th>
<th>EQ_Hypo</th>
<th>EQ_Restitution</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>{contenedor-2}</td>
<td>245</td>
<td>245</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>{contenitore-1}</td>
<td>161</td>
<td>7</td>
<td>6</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>{bak-1}</td>
<td>26</td>
<td>31</td>
<td>11</td>
<td>22</td>
<td>2</td>
</tr>
<tr>
<td>WN15</td>
<td>{container-1}</td>
<td>567</td>
<td>567</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Projection of Reference wordnets to Dutch Source wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Reference Hyponyms</th>
<th>No Match in Source WordNet</th>
<th>Matching WMs in Source</th>
<th>Source WMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>245</td>
<td>209</td>
<td>36</td>
<td>43</td>
</tr>
<tr>
<td>IT</td>
<td>7</td>
<td>2</td>
<td>5</td>
<td>12</td>
</tr>
<tr>
<td>WN15</td>
<td>567</td>
<td>505</td>
<td>62</td>
<td>57</td>
</tr>
</tbody>
</table>

**Comparing projections for the Dutch wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Projected Source WMs</th>
<th>Union</th>
<th>Intersection</th>
<th>Unique in Reference WN</th>
<th>Unique in Source WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>43</td>
<td>55</td>
<td>14</td>
<td>29</td>
<td>12</td>
</tr>
<tr>
<td>IT</td>
<td>12</td>
<td>30</td>
<td>8</td>
<td>4</td>
<td>18</td>
</tr>
<tr>
<td>WN15</td>
<td>57</td>
<td>68</td>
<td>15</td>
<td>42</td>
<td>11</td>
</tr>
<tr>
<td>Union of Reference WNs</td>
<td>59</td>
<td>70</td>
<td>15</td>
<td>44</td>
<td>11</td>
</tr>
</tbody>
</table>

**Errors in Source:**
- wrong translations: 13
- wrong classifications: 0

**Errors in Reference:** 7

**Alternative classifications:**
- voorwerp (object)

**Variant Projection of Unmatched ILI-records from Reference wordnets to Dutch Source Wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Unmatched ILIs</th>
<th>All Senses of ILIs</th>
<th>No Match in Source</th>
<th>Matching ILI in Source</th>
<th>ILIs of Matching Source WM</th>
<th>Intersection with Source Projection</th>
<th>TC-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1046</td>
<td>926</td>
<td>120</td>
<td>184</td>
<td>3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Covering

Major Nodes, hyponyms and equivalence relations

<table>
<thead>
<tr>
<th>Hyperonymic Synsets</th>
<th>No of Hypos All Levels</th>
<th>No of ILIs</th>
<th>EQ_S</th>
<th>EQ_N</th>
<th>EQ_Hypo</th>
<th>EQ_Hypo</th>
<th>EQ_Res</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>{cubierta-1}</td>
<td>425</td>
<td>425</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{cubierta-7}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>{involucro-1}</td>
<td>40</td>
<td>3</td>
<td>1</td>
<td>2</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{copertura-2}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>{bedekking-1}</td>
<td>27</td>
<td>43</td>
<td>3</td>
<td>29</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>{covering-4}</td>
<td>1024</td>
<td>1024</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{covering-5}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Projection of Reference wordnets to Dutch Source wordnet

<table>
<thead>
<tr>
<th>Reference Hyponyms</th>
<th>No Match in Source WordNet</th>
<th>Matching WM in Source</th>
<th>Source WM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>425</td>
<td>330</td>
<td>95</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>113</td>
</tr>
<tr>
<td>IT</td>
<td>3</td>
<td>0</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>WN15</td>
<td>1024</td>
<td>876</td>
<td>148</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>147</td>
</tr>
</tbody>
</table>

Comparing projections for the Dutch wordnet

<table>
<thead>
<tr>
<th>Projected Source WM</th>
<th>Union</th>
<th>Intersection</th>
<th>Unique in Reference WN</th>
<th>Unique in Source WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>113</td>
<td>123</td>
<td>17</td>
<td>96</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>10</td>
</tr>
<tr>
<td>IT</td>
<td>5</td>
<td>27</td>
<td>5</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>22</td>
</tr>
<tr>
<td>WN15</td>
<td>147</td>
<td>156</td>
<td>18</td>
<td>129</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>9</td>
</tr>
</tbody>
</table>

Union of Reference WNs

<table>
<thead>
<tr>
<th>Unmatched ILIs</th>
<th>All Senses of ILIs</th>
<th>No Match in Source</th>
<th>Matching ILI in Source</th>
<th>ILIs Matching Source WM</th>
<th>Intersection with Source Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1846</td>
<td>1615</td>
<td>231</td>
<td>351</td>
<td>10</td>
</tr>
</tbody>
</table>
**High Order Entities**

*Feeling*

**Major Nodes, hyponyms and equivalence relations**

<table>
<thead>
<tr>
<th></th>
<th>Hyperonymic Synsets</th>
<th>No of Hypos All Levels</th>
<th>EQ of Hypos</th>
<th>EQ of ILIs</th>
<th>EQ_Hyper</th>
<th>EQ_Hypo</th>
<th>EQ_Rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>{sentimiento-1}</td>
<td>253</td>
<td>253</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{sensación-6}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{sentir-3} V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{sentir-5} V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>{sentimento-1}</td>
<td>178</td>
<td>32</td>
<td>24</td>
<td>9</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>{percepire-1} V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{provare-7} V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>{voelen-4} V</td>
<td>87</td>
<td>139</td>
<td>19</td>
<td>125</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{voelen-5} V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{gevoel-2}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{gevoel-3}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>{feeling-1}</td>
<td>448</td>
<td>448</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{feeling-6}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{experience-6} V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{feel-7} V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{feel-8} V</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Projection of Reference wordnets to Dutch Source wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Reference Hyponyms</th>
<th>No Match in Source WordNet</th>
<th>Matching WM's in Source</th>
<th>Source WM's</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>253</td>
<td>212</td>
<td>411</td>
<td>56</td>
</tr>
<tr>
<td>IT</td>
<td>32</td>
<td>14</td>
<td>18</td>
<td>38</td>
</tr>
<tr>
<td>WN15</td>
<td>448</td>
<td>398</td>
<td>50</td>
<td>48</td>
</tr>
</tbody>
</table>

**Comparing projections for the Dutch wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Projected Source WMs</th>
<th>Union</th>
<th>Intersection</th>
<th>Unique in Reference WN</th>
<th>Unique in Source WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>56</td>
<td>126</td>
<td>17</td>
<td>39</td>
<td>70</td>
</tr>
<tr>
<td>IT</td>
<td>38</td>
<td>116</td>
<td>9</td>
<td>29</td>
<td>78</td>
</tr>
<tr>
<td>WN15</td>
<td>48</td>
<td>110</td>
<td>25</td>
<td>23</td>
<td>62</td>
</tr>
<tr>
<td>Union of Reference WNs</td>
<td>80</td>
<td>138</td>
<td>29</td>
<td>51</td>
<td>58</td>
</tr>
</tbody>
</table>

**Errors in Source:**
- wrong translations: 14
- wrong classifications: 3

**Errors in Reference:** 2

**Alternative classifications:**
- stimulus: aanvoelen (cause to feel like)
- experience: gewaarwording; ervaring; ervaren; meemaken; ondergaan; gewaarworden; waarnemen;
- attitude: houding; gemoedstoestand; bui/ stemming;
- ability: vermogen

**Variant Projection of Unmatched ILI-records from Reference wordnets to Dutch Source Wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Unmatched ILIs</th>
<th>All Senses of ILIs</th>
<th>No Match in Source</th>
<th>Matching ILI in Source</th>
<th>ILIs Matching Source WM</th>
<th>Intersection with Source Projection</th>
<th>TC-</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>1620</td>
<td>1283</td>
<td>337</td>
<td>449</td>
<td>29</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
## Phenomena

### Major Nodes, hyponyms and equivalence relations

<table>
<thead>
<tr>
<th></th>
<th>Hyperonymic Synsets</th>
<th>No of Hypos All Levels</th>
<th>No of ILIs</th>
<th>EQ_S</th>
<th>EQ_N</th>
<th>EQ_Hypo</th>
<th>EQ_Hyp</th>
<th>EQ_Re</th>
<th>EQ_REST</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>{fenómeno-1}</td>
<td>415</td>
<td>415</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>{caer-57}</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>{fenomeno-1}</td>
<td>100</td>
<td>23</td>
<td>19</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>{verschijnsel-1}</td>
<td>353</td>
<td>241</td>
<td>22</td>
<td>219</td>
<td>3</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>{phenomenon-1}</td>
<td>1012</td>
<td>1012</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Projection of Reference wordnets to Dutch Source wordnet

<table>
<thead>
<tr>
<th></th>
<th>Reference Hyponyms</th>
<th>No Match in Source WordNet</th>
<th>Matching WMs in Source</th>
<th>Source WM</th>
<th>Unique in Reference WN</th>
<th>Unique in Source WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td></td>
<td>415</td>
<td>327</td>
<td>88</td>
<td>102</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td></td>
<td>23</td>
<td>5</td>
<td>18</td>
<td>27</td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td></td>
<td>1012</td>
<td>897</td>
<td>115</td>
<td>118</td>
<td></td>
</tr>
</tbody>
</table>

### Comparing projections for the Dutch wordnet

<table>
<thead>
<tr>
<th></th>
<th>Projected Source WM</th>
<th>Union</th>
<th>Intersection</th>
<th>Unique in Reference WN</th>
<th>Unique in Source WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>102</td>
<td>445</td>
<td>10</td>
<td>92</td>
<td>343</td>
</tr>
<tr>
<td>IT</td>
<td>27</td>
<td>344</td>
<td>6</td>
<td>21</td>
<td>347</td>
</tr>
<tr>
<td>WN15</td>
<td>118</td>
<td>455</td>
<td>16</td>
<td>102</td>
<td>337</td>
</tr>
<tr>
<td>Union of Reference WNs</td>
<td>123</td>
<td>460</td>
<td>16</td>
<td>107</td>
<td>337</td>
</tr>
</tbody>
</table>

### Errors in Source:
- wrong translations: 9
- wrong classifications: 2

### Errors in Reference: 2

### Alternative classifications:
- process/ change/ condition: proces-2; verandering-1; gesteldheid-1 (all more general)
- systems: systeem (mechanisme)
- weather: weersgesteldheid (weather condition)
- power/force: energie-2 -> kracht-6 -> vermogen-; krachtveld
- possibilities: mogelijkheid
- diseases: ziekte-1

### Variant Projection of Unmatched ILI-records from Reference wordnets to Dutch Source Wordnet

<table>
<thead>
<tr>
<th></th>
<th>Unmatched ILIs</th>
<th>All Senses of ILIs</th>
<th>No Match in Source</th>
<th>Matching ILI in Source</th>
<th>ILIs Matching Source WM</th>
<th>Intersection with TC-Source Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total</td>
<td>2591</td>
<td>2119</td>
<td>472</td>
<td>790</td>
<td>17</td>
<td></td>
</tr>
</tbody>
</table>
Appendix Ib Comparing the Spanish wordnet

General Comments:

• Areas compared are:
  • Garment, Places, Furniture, Plants (FirstOrder)
  • Cooking, Sounds (Second Order)
  • Plants can not be properly compared to WN1.5 due to the enormous difference of sets (WN1.5 contains a huge scientific taxonomy). Therefore in this area comparison is only performed using IT and NL as reference

• Major Nodes, hyponyms and equivalence relations (WN1.5 is not considered here)
  • SpWN practically does not have eq relations other than eq_synonym —while NL and IT usually also have other types of equivalence
  • SpWN correspondence to ILI is practically one-to-one, so number of SpWN synsets equals number of ILIs —while NL tends to have more ILIs than synsets; and IT less ILIs than synsets
  • SpWN tends to have more synsets than the others —at this stage of the project

• Projections of Reference WNs to Spanish
  • SpWN tends to already have covered most of the synsets projected from other WNs —except from WN1.5, obviously

• Comparing projections
  • SpWN has few wrong classifications, but an amount of alternative classifications and cases of logical polysemy —to be resolved using complex ILIs or conflating word meanings
  • Both alternative and wrong classifications in SpWN are a reflex of classifications in WN1.5; therefore when such a diagnostic occurs, it applies to both WNs
  • Some cases of wrong translation are found in SpWN.

(Notice that SpWN Synsets including a wrong variant translation are not counted as proper wrong translations, except if this happens to be the only one variant in the synset, since it is considered that the synset has a correct Spanish-ILI correspondence —the refinement work to do in this case is to delete the weird variant.

E.g.: ILI-car, auto ↔ {coche, automóvil, *vagón} no comment
ILI-car, auto ↔ {*vagón} wrong translation)

• Resolving missing synsets
  • Missing synsets are not found using this procedure; the trend is to detect more cases of alternative classification
First Order Entities

Garment

Major Nodes, hyponyms and equivalence relations

<table>
<thead>
<tr>
<th></th>
<th>Hyper Synsets</th>
<th>Nm. of hypos all levels</th>
<th>Nm. of ILIs</th>
<th>EQ_S</th>
<th>EQ_NS</th>
<th>EQ_Hype</th>
<th>EQ_hypo</th>
<th>EQ_rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>indumentaria-1, calzado-1</td>
<td>215</td>
<td>215</td>
<td>215</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IT</td>
<td>indumento 1</td>
<td>156</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>kledingstuk-1</td>
<td>23</td>
<td>36</td>
<td>5</td>
<td>32</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WN15</td>
<td>wear-1, garment-1</td>
<td>277</td>
<td>272</td>
<td>277</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Projection of Reference wordnets to Spanish Source wordnet

<table>
<thead>
<tr>
<th></th>
<th>Reference hypos</th>
<th>No match in SpWN</th>
<th>Matching in SpWN</th>
<th>Synsets in SpWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>215</td>
<td>0</td>
<td>3</td>
<td>128</td>
</tr>
<tr>
<td>IT</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>88</td>
</tr>
<tr>
<td>NL</td>
<td>36</td>
<td>10</td>
<td>26</td>
<td>77</td>
</tr>
<tr>
<td>WN15</td>
<td>277</td>
<td>149</td>
<td>128</td>
<td>128</td>
</tr>
</tbody>
</table>

Comparing projections for the Spanish wordnet

<table>
<thead>
<tr>
<th></th>
<th>Projected Synsets</th>
<th>SpWN Union Classification Intersection</th>
<th>Classification unique in Reference WN</th>
<th>Classification unique in SpWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>215</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IT</td>
<td>3</td>
<td>215</td>
<td>3</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>26</td>
<td>218</td>
<td>23</td>
<td>3</td>
</tr>
<tr>
<td>WN15</td>
<td>128</td>
<td>216</td>
<td>127</td>
<td>1</td>
</tr>
<tr>
<td>Union of Reference WNs</td>
<td>142</td>
<td>219</td>
<td>138</td>
<td>4</td>
</tr>
</tbody>
</table>

Errors in SpWN

Wrong translations: 12
(Some are genuine wrong translations, e.g. 'uplift' -a kind of bra- translated as 'construcción' -the act of building-; other are correct translations of the term in WN1.5, but the choice made by WN1.5 is extremely doubtful so SpWN inherits the error -e.g. WN15:’blue’ as a dress, gloss 'she was wearing blue'; we translate automatically into 'azul')

Errors in Reference WNs

Wrong classifications: 4 (coverings which are not garment, e.g. screen in NL; 'wear' the act as the object in WN1.5; no possible comparison with IT)

Alternative Classifications: no

Variant Projection of Unmatched ILI-records from Reference wordnets to Spanish Source Wordnet

<table>
<thead>
<tr>
<th></th>
<th>No Match in SpWN-Synsets</th>
<th>All Senses of ILIs</th>
<th>No Match in All Senses of SpWN-Synsets</th>
<th>Matching ILI in SpWN</th>
<th>ILIs of SpWN</th>
<th>Matching Intersection with TC-SpWN Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>IT</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>10</td>
<td>29</td>
<td>21</td>
<td>8</td>
<td>8</td>
<td>0</td>
</tr>
<tr>
<td>WN15</td>
<td>149</td>
<td>336</td>
<td>262</td>
<td>74</td>
<td>60</td>
<td>14</td>
</tr>
<tr>
<td>Total</td>
<td>349</td>
<td>273</td>
<td>76</td>
<td>62</td>
<td>62</td>
<td>14</td>
</tr>
</tbody>
</table>

No missing synsets in SpWN are found using this procedure
### Furniture

**Major Nodes, hyponyms and equivalence relations**

<table>
<thead>
<tr>
<th></th>
<th>Hyper Synsets</th>
<th>Num. of hypos all levels</th>
<th>Num. of ILIs</th>
<th>EQ_S</th>
<th>EQ_NS</th>
<th>EQ_Hyper</th>
<th>EQ_hypo</th>
<th>EQ_rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>mobiliario 1</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IT</td>
<td>mobile 2</td>
<td>75</td>
<td>4</td>
<td>65</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>meubelstuk_1</td>
<td>11</td>
<td>15</td>
<td>9</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WN15</td>
<td>furniture</td>
<td>174</td>
<td>174</td>
<td>174</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Projection of Reference wordnets to Spanish Source wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Reference hypos</th>
<th>No match in SpWN</th>
<th>Matching in SpWN</th>
<th>Synsets in SpWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>65</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>75</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>NL</td>
<td>15</td>
<td>6</td>
<td>9</td>
<td>9</td>
</tr>
<tr>
<td>WN15</td>
<td>174</td>
<td>109</td>
<td>65</td>
<td>65</td>
</tr>
</tbody>
</table>

**Comparing projections for the Spanish wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Projected Synsets</th>
<th>SpWN Union</th>
<th>Classification Intersection</th>
<th>Classification unique in Reference WN</th>
<th>Classification unique in SpWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>65</td>
<td>65</td>
<td>4</td>
<td>0</td>
<td>65</td>
</tr>
<tr>
<td>IT</td>
<td>4</td>
<td>65</td>
<td>4</td>
<td>0</td>
<td>61</td>
</tr>
<tr>
<td>NL</td>
<td>9</td>
<td>69</td>
<td>5</td>
<td>4</td>
<td>60</td>
</tr>
<tr>
<td>WN15</td>
<td>65</td>
<td>65</td>
<td>65</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Union of Reference WNs</td>
<td>69</td>
<td>69</td>
<td>65</td>
<td>65</td>
<td>4</td>
</tr>
</tbody>
</table>

**Errors in SpWN:** no

**Errors in Reference WNs:**
- Wrong translations: 1

**Alternative Classifications:** artifact (for 3 unique in NL, which are kinds of small cupboards)

**Variant Projection of Unmatched ILI-records from Reference wordnets to Spanish Source Wordnet**

<table>
<thead>
<tr>
<th></th>
<th>No Match in SpWN-Synsets</th>
<th>All Senses of ILIs</th>
<th>No Match in SpWN-Senses</th>
<th>Matching ILI in SpWN</th>
<th>ILIs of SpWM</th>
<th>Matching SpWM</th>
<th>Intersection with TC-SpWN Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>6</td>
<td>41</td>
<td>26</td>
<td>15</td>
<td>13</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>109</td>
<td>238</td>
<td>193</td>
<td>45</td>
<td>39</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>109</td>
<td>238</td>
<td>193</td>
<td>45</td>
<td>39</td>
<td>6</td>
<td></td>
</tr>
</tbody>
</table>

Some more alternative 'artifact' classifications detected by this procedure
**Places**

**Major Nodes, hyponyms and equivalence relations**

<table>
<thead>
<tr>
<th></th>
<th>Hyper Synsets</th>
<th>Nm. of hypos all levels</th>
<th>Nm. of ILIs</th>
<th>EQ_S</th>
<th>EQ_NS</th>
<th>EQ_Hyper</th>
<th>EQ_hypo</th>
<th>EQ_rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>lugar-1</td>
<td>373</td>
<td>373</td>
<td>373</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IT</td>
<td>luogo1,luogo2</td>
<td>54</td>
<td>26</td>
<td>21</td>
<td>4</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>plaats_1</td>
<td>533</td>
<td>424</td>
<td>95</td>
<td>346</td>
<td>12</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>WN15</td>
<td>location 1</td>
<td>1881</td>
<td>1881</td>
<td>1881</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

**Projection of Reference wordnets to Spanish Source wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Reference hypos</th>
<th>No match in SpWN</th>
<th>Matching in SpWN</th>
<th>Synsets in SpWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>373</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>26</td>
<td>1</td>
<td>25</td>
<td>25</td>
</tr>
<tr>
<td>NL</td>
<td>424</td>
<td>127</td>
<td>297</td>
<td>298</td>
</tr>
<tr>
<td>WN15</td>
<td>1881</td>
<td>1508</td>
<td>373</td>
<td>373</td>
</tr>
</tbody>
</table>

**Comparing projections for the Spanish wordnet**

<table>
<thead>
<tr>
<th></th>
<th>Projected Synsets</th>
<th>SpWN Union</th>
<th>Classification</th>
<th>Classification unique in Reference WN</th>
<th>Classification unique in SpWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>373</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>25</td>
<td>382</td>
<td>16</td>
<td>9</td>
<td>357</td>
</tr>
<tr>
<td>NL</td>
<td>298</td>
<td>613</td>
<td>58</td>
<td>240</td>
<td>315</td>
</tr>
<tr>
<td>WN15</td>
<td>373</td>
<td>373</td>
<td>373</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Union of Reference WNs</td>
<td>619</td>
<td>619</td>
<td>373</td>
<td>246</td>
<td>0</td>
</tr>
</tbody>
</table>

**Errors in SpWN**
Wrong translations: 15
Wrong Classification: 1 (way/road as artifact)

**Errors in Reference WNs**
Wrong translations: 15
Wrong classification: 17 (mainly anatonical terminology: 'callosity', 'tuberosity' as places; also other as 'rubbish' as place; 'plant' as place; 'opening' as place)
Doubtful classifications: container as a place - SpWN hasn't got the 245 hyponyms of 'container' classified as places, while others have -; rack/stand as a place

**Alternative Classifications**: cognition (for imaginary places); geographyc terms → land → object (e.g. 'depression', 'island', 'tundra', 'peninsula'); building → artifact (e.g. 'office'); facility/installation (e.g. sports fields)
Notice constructions and installations are clear cases of logical polysemny.

**Variant Projection of Unmatched ILI-records from Reference wordnets to Spanish Source Wordnet**

<table>
<thead>
<tr>
<th></th>
<th>No Match in SpWN-Synsets</th>
<th>All Senses of ILIs</th>
<th>No Match in SpWN-Senses</th>
<th>All Matching ILI in SpWN</th>
<th>ILIs of Matching Source WM</th>
<th>Intersection with TC-Project SpWN Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>1</td>
<td>33</td>
<td>15</td>
<td>18</td>
<td>13</td>
<td>5</td>
</tr>
<tr>
<td>IT</td>
<td>127</td>
<td>636</td>
<td>433</td>
<td>203</td>
<td>182</td>
<td>21</td>
</tr>
<tr>
<td>NL</td>
<td>1508</td>
<td>2896</td>
<td>2423</td>
<td>473</td>
<td>422</td>
<td>51</td>
</tr>
<tr>
<td>Total</td>
<td>3303</td>
<td>2714</td>
<td>589</td>
<td>535</td>
<td>54</td>
<td></td>
</tr>
</tbody>
</table>

Many cases of alternative classifications noticed before (containers, some imaginary spaces,...) appear here as missing - not classified as places in SpWN
Plants

Only first 3 levels of WN1.5 are used for comparison

Major Nodes, hyponyms and equivalence relations

<table>
<thead>
<tr>
<th></th>
<th>Hyper Synsets</th>
<th>Nm. of hypos all levels</th>
<th>Nm. of ILIs</th>
<th>EQ_S</th>
<th>EQ_NS</th>
<th>EQ_Hyper</th>
<th>EQ_hypo</th>
<th>EQ_rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>plantal</td>
<td>467</td>
<td>467</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>IT</td>
<td>piantal</td>
<td>474</td>
<td>261</td>
<td>253</td>
<td>6</td>
<td>10</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>plant1 gewas1</td>
<td>28</td>
<td>40</td>
<td>7</td>
<td>29</td>
<td>0</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>WN15</td>
<td>plant1 (first 3 levels)</td>
<td>802</td>
<td>802</td>
<td>802</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Projection of Reference wordnets to Spanish Source wordnet

<table>
<thead>
<tr>
<th></th>
<th>Reference hypos</th>
<th>No match in SpWN</th>
<th>Matching in SpWN</th>
<th>Synsets in SpWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>467</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>261</td>
<td>86</td>
<td>175</td>
<td>175</td>
</tr>
<tr>
<td>NL</td>
<td>40</td>
<td>7</td>
<td>33</td>
<td>33</td>
</tr>
<tr>
<td>WN15</td>
<td>802</td>
<td>670</td>
<td>132</td>
<td>132</td>
</tr>
</tbody>
</table>

Comparing projections for the Spanish wordnet

<table>
<thead>
<tr>
<th></th>
<th>Projected Synsets</th>
<th>SpWN</th>
<th>Union</th>
<th>Classification Intersection</th>
<th>Classification unique in Reference WN</th>
<th>Classification unique in SpWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>175</td>
<td>472</td>
<td>170</td>
<td>5</td>
<td>297</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>33</td>
<td>481</td>
<td>19</td>
<td>14</td>
<td>448</td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>132</td>
<td>467</td>
<td>132</td>
<td>0</td>
<td>335</td>
<td></td>
</tr>
<tr>
<td>Union of Reference WNs</td>
<td>272</td>
<td>486</td>
<td>253</td>
<td>19</td>
<td>214</td>
<td></td>
</tr>
</tbody>
</table>

Errors in SpWN: no

Errors in Reference WNs:
- Wrong classification: 5
- Wrong translation: 3

Alternative Classifications: microorganism (for 'alga'); vegetables (for edible roots and seeds)

Variant Projection of Unmatched ILI-records from Reference wordnets to Spanish Source Wordnet

<table>
<thead>
<tr>
<th></th>
<th>No Match in SpWN-Synsets</th>
<th>All Senses of ILIs</th>
<th>No Match in SpWN-Senses</th>
<th>Matching ILI in SpWN</th>
<th>ILIs of SpWM</th>
<th>Matching ILIs of SpWM</th>
<th>Intersection with TC-SpWN Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>86</td>
<td>253</td>
<td>201</td>
<td>52</td>
<td>46</td>
<td>6</td>
<td>23</td>
</tr>
<tr>
<td>IT</td>
<td>7</td>
<td>37</td>
<td>24</td>
<td>13</td>
<td>9</td>
<td>4</td>
<td>13</td>
</tr>
<tr>
<td>NL</td>
<td>670</td>
<td>925</td>
<td>870</td>
<td>55</td>
<td>42</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>1158</td>
<td>1046</td>
<td>112</td>
<td>89</td>
<td>23</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

More cases of Alternative classification -Vegetables- found; also alternative classification: Fruits (kiwi, peanut), found
Wrong classification in SpWN found: Mistletoe as parasite but not as plant.
Higher Order entities

Sounds

Major Nodes, hyponyms and equivalence relations

<table>
<thead>
<tr>
<th></th>
<th>Hyper Synsets</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>sonido-2, SONAR-3, emitir_sonidos-1</td>
<td>139</td>
<td>139</td>
<td>139</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IT</td>
<td>{rumore 1, suono 1} {emettere 3, produrre 5}</td>
<td>45</td>
<td>41</td>
<td>21</td>
<td>21</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>{geluid_2n}, {klinken_2v}</td>
<td>22</td>
<td>33</td>
<td>9</td>
<td>24</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WN15</td>
<td>sound_13v, sound_5n, utter-3v</td>
<td>271</td>
<td>271</td>
<td>271</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Projection of Reference wordnets to Spanish Source wordnet

<table>
<thead>
<tr>
<th></th>
<th>Reference hypos</th>
<th>No match in SpWN</th>
<th>Matching in SpWN</th>
<th>Synsets in SpWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>139</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>IT</td>
<td>41</td>
<td>12</td>
<td>29</td>
<td>29</td>
</tr>
<tr>
<td>NL</td>
<td>33</td>
<td>5</td>
<td>28</td>
<td>28</td>
</tr>
<tr>
<td>WN15</td>
<td>271</td>
<td>132</td>
<td>139</td>
<td>139</td>
</tr>
</tbody>
</table>

Comparing projections for the Spanish wordnet

<table>
<thead>
<tr>
<th></th>
<th>Projected Synsets</th>
<th>SpWN</th>
<th>Union</th>
<th>Classification Intersection</th>
<th>Classification unique in Reference WN</th>
<th>Classification unique in SpWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>139</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>29</td>
<td>149</td>
<td>19</td>
<td>10</td>
<td>120</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>28</td>
<td>151</td>
<td>16</td>
<td>12</td>
<td>123</td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>139</td>
<td>139</td>
<td>139</td>
<td>0</td>
<td>0</td>
<td></td>
</tr>
</tbody>
</table>

Union of Reference WNs

|  | 160 | 160 | 139 | 21 | 0 |

Errors in SpWN

Wrong Classification: 2

Errors in Reference WNs

Wrong classification: 3
Wrong translation: 4

Alternative Classifications: communicate, breathe

Variant Projection of Unmatched ILI-records from Reference wordnets to Spanish Source Wordnet

<table>
<thead>
<tr>
<th></th>
<th>No Match in SpWN-Synsets</th>
<th>All Senses of ILIs</th>
<th>No Match in All Senses</th>
<th>Matching ILI in SpWN</th>
<th>ILIs of SpWM</th>
<th>Matching</th>
<th>Intersection with TC-SpWN Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>29</td>
<td>145</td>
<td>92</td>
<td>53</td>
<td>40</td>
<td>13</td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>28</td>
<td>52</td>
<td>38</td>
<td>14</td>
<td>12</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>139</td>
<td>682</td>
<td>415</td>
<td>267</td>
<td>208</td>
<td>59</td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>767</td>
<td>477</td>
<td>290</td>
<td>227</td>
<td>63</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Another alternative classification detected: music
Cooking

Major Nodes, hyponyms and equivalence relations

<table>
<thead>
<tr>
<th></th>
<th>Hyper Synsets</th>
<th>Nm. hypos levels</th>
<th>Nm. of all ILIs</th>
<th>Eq_S</th>
<th>Eq_NS</th>
<th>Eq_Hyper</th>
<th>Eq_hypo</th>
<th>Eq_rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>cocer-3v, cocina-1n</td>
<td>20</td>
<td>20</td>
<td>20</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IT</td>
<td>cuocere 1v, preparare 3v, cuocere 2v</td>
<td>24</td>
<td>13</td>
<td>8</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>klaarmaken_2v, koken_2v</td>
<td>3</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WN15</td>
<td>cook1v,cook-2v, cook-3v, cook-4v, cooking-1n</td>
<td>57</td>
<td>57</td>
<td>57</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Projection of Reference wordnets to Spanish Source wordnet

<table>
<thead>
<tr>
<th></th>
<th>Reference hypos</th>
<th>No match in SpWN</th>
<th>Matching in SpWN</th>
<th>Synsets in SpWN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>20</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>13</td>
<td>2</td>
<td>11</td>
<td>11</td>
</tr>
<tr>
<td>NL</td>
<td>6</td>
<td>2</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>WN15</td>
<td>57</td>
<td>34</td>
<td>23</td>
<td>23</td>
</tr>
</tbody>
</table>

Comparing projections for the Spanish wordnet

|       | Projected Synsets | SpWN Union | Classification Intersection | Classification unique in Reference WN | Classification unique in SpWN | |
|-------|-------------------|------------|-----------------------------|---------------------------------------|-------------------------------|
| ES    | 20                |            |                             |                                       |                               |
| IT    | 11                | 24         | 7                           | 4                                     | 13                            |
| NL    | 4                 | 23         | 1                           | 3                                     | 19                            |
| WN15  | 23                | 23         | 20                          | 3                                     | 0                             |
| Union of Reference WNs | 26     | 26         | 20                          | 6                                     | 0                             |

Errors in SpWN: no
Errors in Reference WNs: no
Alternative Classifications: creation (for the act of cooking), change (for 'caramelize')

Variant Projection of Unmatched ILI-records from Reference wordnets to Spanish Source Wordnet

<table>
<thead>
<tr>
<th></th>
<th>No Match in SpWN-Synsets</th>
<th>All Senses of ILIs</th>
<th>No Match in All Senses</th>
<th>Matching ILI in SpWN</th>
<th>ILIs of Matching ILI in SpWM</th>
<th>Intersection with TC-SpWN Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>2</td>
<td>3</td>
<td>3</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>WN15</td>
<td>34</td>
<td>103</td>
<td>80</td>
<td>23</td>
<td>22</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>103</td>
<td>80</td>
<td>23</td>
<td>23</td>
<td>1</td>
</tr>
</tbody>
</table>

Another Alternative Classification found: creation
Appendix Ic Comparing the Italian wordnet

First Order Entities

Animal

Hyponyms for “animal 1” with the type of equivalences

<table>
<thead>
<tr>
<th></th>
<th>Hyperonymic Synsets</th>
<th>No Hypos of All Levels</th>
<th>No ILIs</th>
<th>EQ_S</th>
<th>EQ_NS</th>
<th>EQ_Hyper</th>
<th>EQ_Hypo</th>
<th>EQ_Rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>{animal-1}</td>
<td>682</td>
<td>682</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IT</td>
<td>{animale 1, bestia 1, organismo animale 20}</td>
<td>563</td>
<td>318</td>
<td>302</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>{dier_1, gedierte_1}</td>
<td>26</td>
<td>43</td>
<td>13</td>
<td>23</td>
<td>0</td>
<td>9</td>
<td>0</td>
</tr>
<tr>
<td>WN15*</td>
<td>{animal, animate being, beast, brute, creature, fauna}</td>
<td>2017</td>
<td>2017</td>
<td>2017</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Projection of TC-Reference projection on Source wordnet

<table>
<thead>
<tr>
<th></th>
<th>Reference Hyponyms</th>
<th>No Match in Source WordNet</th>
<th>Matching WMs in Source</th>
<th>Source WMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>681</td>
<td>473</td>
<td>209</td>
<td>682</td>
</tr>
<tr>
<td>IT</td>
<td>563</td>
<td>0</td>
<td>318</td>
<td>318</td>
</tr>
<tr>
<td>NL</td>
<td>26</td>
<td>17</td>
<td>28</td>
<td>43</td>
</tr>
<tr>
<td>WN15*</td>
<td>2017</td>
<td>1794</td>
<td>224</td>
<td>2017</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Projected Source WMs</th>
<th>Union</th>
<th>Intersection</th>
<th>Unique in Reference WN</th>
<th>Unique in Source WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>363</td>
<td>566</td>
<td>361</td>
<td>2</td>
<td>203</td>
</tr>
<tr>
<td>IT</td>
<td>495</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>157</td>
<td>565</td>
<td>156</td>
<td>1</td>
<td>408</td>
</tr>
<tr>
<td>WN15*</td>
<td>310</td>
<td>567</td>
<td>308</td>
<td>2</td>
<td>257</td>
</tr>
</tbody>
</table>

Union of Reference WNs

Errors in source: 1

Variant Projection of Unmatched ILI-records from Reference wordnets to Source Wordnet

<table>
<thead>
<tr>
<th></th>
<th>Unique Reference</th>
<th>All Senses of ILIs</th>
<th>No Match in Source</th>
<th>Matching ILI in Source</th>
<th>ILIs of Matching Source WM</th>
<th>Intersection with TC-Source Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>473</td>
<td>892</td>
<td>844</td>
<td>48</td>
<td>23</td>
<td>45</td>
</tr>
<tr>
<td>IT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>15</td>
<td>13</td>
<td>10</td>
<td>3</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>WN15</td>
<td>1794*</td>
<td>2315</td>
<td>2246</td>
<td>69</td>
<td>62</td>
<td>38</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* The comparison to WN1.5 has been performed on a smaller set of hyponyms, comprehensive of the first level below {animal, animate being, beast, brute, creature, fauna} and of the whole subsets of {bird} and {mammal}, via {chordate 1} and {vertebrate 1}. 
### Artist

Hyponyms for “**artist**” with the type of equivalences

<table>
<thead>
<tr>
<th></th>
<th>Hypernymic Synsets</th>
<th>No Hypos All Levels</th>
<th>No of ILIs</th>
<th>EQ_S</th>
<th>EQ_NS</th>
<th>EQ_Hyper</th>
<th>EQ_Hypo</th>
<th>EQ_Rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>{artista 2, pintor1}</td>
<td>30</td>
<td>30</td>
<td>30</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IT</td>
<td>{artista 1}</td>
<td>91</td>
<td>38</td>
<td>33</td>
<td>7</td>
<td>3</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>{kunstenaar 1}</td>
<td>4</td>
<td>4</td>
<td>3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td></td>
<td>{artiest 1}</td>
<td>71</td>
<td>71</td>
<td>71</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WN15</td>
<td>{artist 1}</td>
<td>71</td>
<td>71</td>
<td>71</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Projection of TC-Reference projection on Source wordnet

<table>
<thead>
<tr>
<th></th>
<th>Reference Hyponyms</th>
<th>No Match in Source WordNet</th>
<th>Matching WM in Source</th>
<th>Source WM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>30</td>
<td>25</td>
<td>6</td>
<td>30</td>
</tr>
<tr>
<td>IT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>4</td>
<td>0</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>WN15</td>
<td>71</td>
<td>59</td>
<td>13</td>
<td>71</td>
</tr>
<tr>
<td>Total</td>
<td>105</td>
<td>84</td>
<td>23</td>
<td>105</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Projected Source WM</th>
<th>Union</th>
<th>Intersection</th>
<th>Unique in Reference WN</th>
<th>Unique in Source WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>6</td>
<td>96</td>
<td>2</td>
<td>4</td>
<td>92</td>
</tr>
<tr>
<td>IT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>11</td>
<td>92</td>
<td>11</td>
<td>0</td>
<td>81</td>
</tr>
<tr>
<td>WN15</td>
<td>15</td>
<td>104</td>
<td>3</td>
<td>12</td>
<td>89</td>
</tr>
<tr>
<td>Union of Reference WNs</td>
<td>25</td>
<td>104</td>
<td>13</td>
<td>12</td>
<td>79</td>
</tr>
</tbody>
</table>

Errors in source: 0

Variant Projection of Unmatched ILI-records from Reference wordnets to Source Wordnet

<table>
<thead>
<tr>
<th></th>
<th>Unique Reference</th>
<th>All Senses of ILIs</th>
<th>No Match in Source</th>
<th>Matching ILI Source</th>
<th>ILIs Matching Source WM</th>
<th>Intersection with TC-Source Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>25</td>
<td>58</td>
<td>53</td>
<td>5</td>
<td>39</td>
<td>56</td>
</tr>
<tr>
<td>IT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>4</td>
<td>0</td>
<td>53</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WN15</td>
<td>71</td>
<td>112</td>
<td>104</td>
<td>8</td>
<td>11</td>
<td>9</td>
</tr>
<tr>
<td>Total</td>
<td>100</td>
<td>170</td>
<td>157</td>
<td>13</td>
<td>50</td>
<td>65</td>
</tr>
</tbody>
</table>
### Worker

Hyponyms for “worker 2” with the type of equivalences

<table>
<thead>
<tr>
<th></th>
<th>Hyperonymic Synsets</th>
<th>No Hypos of ILIs</th>
<th>No ILIs</th>
<th>$EQ_S$</th>
<th>$EQ_NS$</th>
<th>$EQ_Hyper$</th>
<th>$EQ_Hypo$</th>
<th>$EQ_Rest$</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>trabajador 1</td>
<td>356</td>
<td>356</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT</td>
<td>lavoratore 1</td>
<td>552</td>
<td>251</td>
<td>204</td>
<td>55</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>werknemer 1</td>
<td>9</td>
<td>11</td>
<td>5</td>
<td>7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>worker 2</td>
<td>675</td>
<td>675</td>
<td>675</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Projection of TC-Reference projection on Source wordnet

<table>
<thead>
<tr>
<th></th>
<th>Reference Hyponyms</th>
<th>No Match in Source WordNet</th>
<th>Matching WMs in Source</th>
<th>Source WMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>356</td>
<td>229</td>
<td>128</td>
<td>356</td>
</tr>
<tr>
<td>IT</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>9</td>
<td>4</td>
<td>7</td>
<td>11</td>
</tr>
<tr>
<td>WN15</td>
<td>675</td>
<td>523</td>
<td>153</td>
<td>675</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Errors in source: 2

Variant Projection of Unmatched ILLI-records from Reference wordnets to Source Wordnet

<table>
<thead>
<tr>
<th></th>
<th>Unique Reference</th>
<th>All Senses of ILIs</th>
<th>No Match in Source ILI in Source</th>
<th>Matching ILI in Source</th>
<th>ILIs Matching Source WM</th>
<th>Intersection with TC-Source Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>229</td>
<td>444</td>
<td>403</td>
<td>41</td>
<td>95</td>
<td>16</td>
</tr>
<tr>
<td>IT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>4</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WN15</td>
<td>523</td>
<td>917</td>
<td>849</td>
<td>68</td>
<td>122</td>
<td>35</td>
</tr>
<tr>
<td>Total</td>
<td>756</td>
<td>1362</td>
<td>1253</td>
<td>109</td>
<td>217</td>
<td>51</td>
</tr>
</tbody>
</table>
**Instrument**

Hyponyms for “instrument 2” with the type of equivalences

<table>
<thead>
<tr>
<th></th>
<th>Hyperonymic Synsets</th>
<th>No of Hypos All Levels</th>
<th>EQ_S</th>
<th>EQ_NS</th>
<th>EQ_Hyper</th>
<th>EQ_Hypo</th>
<th>EQ_Rest</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ES</strong></td>
<td>{ herramienta-1, instrumento-3}</td>
<td>185</td>
<td>185</td>
<td>185</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td><strong>IT</strong></td>
<td>{strumento 1, attrezzo 1, arnese 1, utensile 1}</td>
<td>867</td>
<td>393</td>
<td>330</td>
<td>44</td>
<td>58</td>
<td>0</td>
</tr>
<tr>
<td><strong>NL</strong></td>
<td>{ instrument_1}</td>
<td>437</td>
<td>266</td>
<td>72</td>
<td>199</td>
<td>31</td>
<td>1</td>
</tr>
<tr>
<td><strong>WN15</strong></td>
<td>{ instrument 2}</td>
<td>509</td>
<td>509</td>
<td>509</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Reference Hyponyms</th>
<th>No Match in Source WordNet</th>
<th>Matching WMs in Source</th>
<th>Source WMs</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ES</strong></td>
<td>185</td>
<td>101</td>
<td>85</td>
<td>186</td>
</tr>
<tr>
<td><strong>IT</strong></td>
<td>867</td>
<td>0</td>
<td>393</td>
<td>393</td>
</tr>
<tr>
<td><strong>NL</strong></td>
<td>437</td>
<td>163</td>
<td>103</td>
<td>266</td>
</tr>
<tr>
<td><strong>WN15</strong></td>
<td>509</td>
<td>379</td>
<td>131</td>
<td>509</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Source</th>
<th>Projected Source WMs</th>
<th>Union</th>
<th>Intersection</th>
<th>Unique in Reference WN</th>
<th>Unique in Source WN</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ES</strong></td>
<td>163</td>
<td>1002</td>
<td>29</td>
<td>134</td>
<td>839</td>
</tr>
<tr>
<td><strong>IT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NL</strong></td>
<td>180</td>
<td>1023</td>
<td>25</td>
<td>155</td>
<td>843</td>
</tr>
<tr>
<td><strong>WN15</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Alternative Classifications: for Italian, container has hyperonym Artifact, not Instrument

Variant Projection of Unmatched ILI-records from Reference wordnets to Source Wordnet

<table>
<thead>
<tr>
<th>Source</th>
<th>Unique Reference</th>
<th>All Senses of ILIs</th>
<th>No Match in Source ILIs</th>
<th>Matching ILI in Source</th>
<th>ILIs of Matching Source WM</th>
<th>Intersection with Source Projection</th>
<th>TC-Source Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>ES</strong></td>
<td>101</td>
<td>261</td>
<td>240</td>
<td>21</td>
<td>19</td>
<td>6</td>
<td></td>
</tr>
<tr>
<td><strong>IT</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>NL</strong></td>
<td>163</td>
<td>344</td>
<td>322</td>
<td>22</td>
<td>23</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td><strong>WN15</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
### Vehicle

Hyponyms for “vehicle 1” with the type of equivalences

<table>
<thead>
<tr>
<th></th>
<th>Hyperonymic Synsets</th>
<th>No of Hypos All Levels</th>
<th>No of ILIs</th>
<th>EQ_S</th>
<th>EQ_NS</th>
<th>EQ_Hyper</th>
<th>EQ_Hypo</th>
<th>EQ_Rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>{ transporte-5}</td>
<td>189</td>
<td>189</td>
<td>189</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IT</td>
<td>{ veicolo 1}</td>
<td>172</td>
<td>72</td>
<td>60</td>
<td>6</td>
<td>15</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>{ voertuig_1}</td>
<td>21</td>
<td>35</td>
<td>7</td>
<td>27</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>WN15</td>
<td>{ vehicle 1}</td>
<td>410</td>
<td>410</td>
<td>410</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Reference Hyponyms</th>
<th>No Match in Source WordNet</th>
<th>Matching WM in Source</th>
<th>Source WM</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>189</td>
<td>134</td>
<td>56</td>
<td>190</td>
</tr>
<tr>
<td>IT</td>
<td>172</td>
<td>0</td>
<td>72</td>
<td>72</td>
</tr>
<tr>
<td>NL</td>
<td>21</td>
<td>23</td>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>WN15</td>
<td>410</td>
<td>343</td>
<td>68</td>
<td>410</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th></th>
<th>Projected Source WMs</th>
<th>Union</th>
<th>Intersection</th>
<th>Unique in Reference WN</th>
<th>Unique in Source WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>97</td>
<td>176</td>
<td>94</td>
<td>3</td>
<td>79</td>
</tr>
<tr>
<td>IT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>59</td>
<td>217</td>
<td>15</td>
<td>44</td>
<td>158</td>
</tr>
<tr>
<td>WN15</td>
<td>110</td>
<td>177</td>
<td>106</td>
<td>4</td>
<td>67</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Errors in source:** *spartineve* is a mechanical device and not a vehicle: error of mapping.

#### Variant Projection of Unmatched ILL-records from Reference wordnets to Source Wordnet

<table>
<thead>
<tr>
<th></th>
<th>Unique Reference</th>
<th>All Senses of ILLs</th>
<th>No Match in Source</th>
<th>Matching ILL in Source</th>
<th>ILIs of Matching Source WM</th>
<th>Intersection with TC-Source Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>134</td>
<td>355</td>
<td>344</td>
<td>11</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>IT</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>23</td>
<td>68</td>
<td>66</td>
<td>2</td>
<td>4</td>
<td>0</td>
</tr>
<tr>
<td>WN15</td>
<td>343</td>
<td>681</td>
<td>658</td>
<td>23</td>
<td>15</td>
<td>11</td>
</tr>
<tr>
<td>Total</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Higher Order entities

Movement

Major Nodes, hyponyms and equivalence relations

<table>
<thead>
<tr>
<th></th>
<th>Hyper Synsets</th>
<th>Nm. of hypos all levels</th>
<th>Nm. Of ILIs</th>
<th>EQ_S</th>
<th>EQ_NS</th>
<th>EQ_hyper</th>
<th>EQ_hypo</th>
<th>EQ_rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>movimiento 8, movimiento 2, movimiento 1, mover 1, mover 3, move 3, move 4</td>
<td>600</td>
<td>600</td>
<td>600</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>IT</td>
<td>movimento 1, muoversi 1, muovere 1</td>
<td>148</td>
<td>98</td>
<td>51</td>
<td>57</td>
<td>4</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>NL</td>
<td>beweging 1, bewegen 1, bewegen 2</td>
<td>1313</td>
<td>1304</td>
<td>94</td>
<td>1255</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>WN15</td>
<td>motion 1, motion 2, motion 5, move 1, move 4, move 2</td>
<td>1891</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Projections of TC-Reference on It WN

<table>
<thead>
<tr>
<th></th>
<th>Reference hypos</th>
<th>No match in It WN</th>
<th>Matching in It WN</th>
<th>It WN WMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>600</td>
<td>553</td>
<td>47</td>
<td>69</td>
</tr>
<tr>
<td>IT</td>
<td>145</td>
<td>-</td>
<td>-</td>
<td>130</td>
</tr>
<tr>
<td>NL</td>
<td>1313</td>
<td>1200</td>
<td>104</td>
<td>142</td>
</tr>
<tr>
<td>WN15</td>
<td>1891</td>
<td>1817</td>
<td>74</td>
<td>102</td>
</tr>
</tbody>
</table>

Comparing projections

<table>
<thead>
<tr>
<th></th>
<th>Projected Synsets</th>
<th>It WN</th>
<th>Union</th>
<th>Classification Intersection</th>
<th>Classification unique Reference WN</th>
<th>Classification unique in It WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>69</td>
<td>153</td>
<td>64</td>
<td>5</td>
<td>84</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>142</td>
<td>200</td>
<td>90</td>
<td>52</td>
<td>58</td>
<td></td>
</tr>
<tr>
<td>WN15</td>
<td>102</td>
<td>155</td>
<td>95</td>
<td>7</td>
<td>53</td>
<td></td>
</tr>
<tr>
<td>Union of Reference WNs</td>
<td>164</td>
<td>203</td>
<td>109</td>
<td>55</td>
<td>39</td>
<td></td>
</tr>
</tbody>
</table>

Possibly wrong classification in It Wn: 4
Example errors in source: 1 (schema di gioco -play)

Alternative Classifications wrt the other Wns: 51. Examples alternative classifications:
- ‘battuta’ (-sport- the act of swinging or striking at a ball...) is a hyponym of ‘azione’ (act, action) via ‘colpo’ (the act of hitting);
- in Dutch some natural phenomena (like storm, shower ecc..) are hyponyms of ‘movement’, while, in Italian, they are ‘atmospheric phenomena’;
- ‘play’ (a preset plan of action) is classified as hyponym of ‘movement’;
- ‘rabbrividire’ (feel shivers because of cold, fear, etc.) is classified as a perception in Italian.

Possible Missing Synsets

<table>
<thead>
<tr>
<th></th>
<th>No Match in It WN Synsets</th>
<th>All Senses of ILIs</th>
<th>No Match in All Senses</th>
<th>Matching ILI in It WN</th>
<th>ILIs of Matching It WM</th>
<th>Intersection with TC-It WN Projection</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>553</td>
<td>3925</td>
<td>3751</td>
<td>174</td>
<td>328</td>
<td>70</td>
</tr>
<tr>
<td>NL</td>
<td>1200</td>
<td>6878</td>
<td>6563</td>
<td>315</td>
<td>424</td>
<td>86</td>
</tr>
<tr>
<td>WN15</td>
<td>1817</td>
<td>7795</td>
<td>7438</td>
<td>357</td>
<td>592</td>
<td>89</td>
</tr>
</tbody>
</table>

Part of the synsets still need to be linked to ILI. Most problems seem however due to different classifications in the various Wns. A few cases can be reconsidered for a different classification in the It Wn.
**Knowledge**

**Major Nodes, hyponyms and equivalence relations**

<table>
<thead>
<tr>
<th>Hyper Synsets</th>
<th>Nm. hypos of all</th>
<th>Of</th>
<th>EQ_S</th>
<th>EQ_NS</th>
<th>EQ_hyper</th>
<th>EQ_hypo</th>
<th>EQ_rest</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES Información 1, pensamiento 2, teoría 3, disciplina 2, pensamiento 1</td>
<td>159</td>
<td>159</td>
<td>159</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IT Conoscenza 3, disciplina 1, conoscere 1</td>
<td>223</td>
<td>15</td>
<td>13</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NL Kennis 2, Weten 2</td>
<td>53</td>
<td>69</td>
<td>20</td>
<td>53</td>
<td>1</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Projections of TC-Reference on It WN**

<table>
<thead>
<tr>
<th>Reference hypos</th>
<th>No match in It WN</th>
<th>Matching in It WN</th>
<th>It WN WMNs</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>159</td>
<td>143</td>
<td>17</td>
</tr>
<tr>
<td>IT</td>
<td>223</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>NL</td>
<td>53</td>
<td>53</td>
<td>16</td>
</tr>
</tbody>
</table>

**Comparing projections**

<table>
<thead>
<tr>
<th>Projected Synsets</th>
<th>It WN</th>
<th>Union</th>
<th>Classification Intersection</th>
<th>Classification unique in Reference WN</th>
<th>Classification unique in It WN</th>
</tr>
</thead>
<tbody>
<tr>
<td>ES</td>
<td>20</td>
<td>240</td>
<td>20</td>
<td>220</td>
<td></td>
</tr>
<tr>
<td>NL</td>
<td>51</td>
<td>235</td>
<td>11</td>
<td>219</td>
<td></td>
</tr>
</tbody>
</table>

Most of the Italian concepts still need to be linked to the ILI.

**Alternative Classifications wrt the other WNs.** Example alternative classifications: ‘Record’ (a document that can serve as legal evidence) in Dutch WordNet is a hyponym of information while in Italian Wordnet is a hyponym of textual matter.
Appendix II Software utilities for graph-comparison

We include here a list of programs that have been used to obtain all the statistical data for the overall comparison in this report. The list appears in alphabetical order. All this software can be found in the ftp-site of the project, as “/tools/wns_compare.tar.gz”.

- ‘cadenes.pl’: the aim of this program is to write to the standard output a list of the chain lengths with the number of occurrences of each length. This program needs in standard input a file of chains.

  Syntax:
  ```bash
  cadenes.pl < <in_file> > <out_file>
  ```

  Example:
  Query:
  ```bash
  cadenes.pl < example.txt
  ```

  Input:
  ```bash
  00002728 00004865 00621770 00002728 00004865 02766721 00002728 00004865 03207851 00002728 00004865 05839075 06193747 00002728 00004865 05842570 00002728 00004865 05843454 00002728 00004865 05844200 05963844
  ```

  Output:
  ```bash
  3: 5
  4: 2
  ```

- ‘chains.pl’: the aim of this program is to query a file of chains created with the ‘graf.pl’ program which is described later. For all queries there are two consult modes, the single mode (only for counting occurrences) and the verbose mode (for counting and extracting occurrences). The chains have the next format:

  ```bash
  ...
  ```

  where the codes for describing coverage of nodes, (****), and edges, [****] consits of tuples of 4 elements, 1 if the corresponding language covers the node/edge or 0 in the other case.

  The queries we can perform using this program are:

  - **Complete Node Chains**: Giving a selection of languages, $L$, this query obtains the list of complete chains node-covered by all the languages in $L$.

    Syntax:
    ```bash
    chains.pl [-v] “cv(<languages>)” < <in_file> > <out_file>
    ```

    where:
    ```bash
    <languages> is: a sequence of the languages (using letters) we want to consult. For example: ed means English and Dutch.
    [-v]: if this optional parameters is present, the program queries in verbose mode.
    <in_file>: file containing the chains
    <out_file>: file where the results will be placed
    ```
Example:

Query:
chains.pl -v "cv(s)" < example.chains

for getting all chains in example.chains completely node-covered for Spanish.

Input:

00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1000]
05636823(1000){mental_energy} [1000] 05637150(1000){libidinal_energy} [1000]
05637285(1000){cathexis}

00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1000]
05636823(1000){mental_energy} [1000] 05636964(1100){incitement} [1100]
05637094(1100){goad}

00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100]
05633277(1111){life}

00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100]
05635834(1100){ethical_motive} [1100] 05636022(1100){hedonism}

00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100]
05635834(1100){ethical_motive} [1100] 05636133(1100){conscience} [1100]
05636402(1000){small_voice}

00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100]
05635834(1100){ethical_motive} [1100] 05636133(1100){conscience} [1100]
03839123(1100){superego}

00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100]
05635834(1100){ethical_motive} [1100] 05636665(1000){christ_within}

00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1000]
05634773(1000){irrational_motive} 05635682(1000){compulsion}

00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1000]
05634773(1000){irrational_motive} [1000] 05635349(1100){mania} [1000]
05635472(1000){monomania}

Output:

00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100]
05633277(1111){life}

00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100]
05635834(1100){ethical_motive} [1100] 05636022(1100){hedonism}

4 chains.

- **Complete Edge Chains**: Giving a selection of languages, $L$, this query obtains the list of complete chains edge-covered by all the languages in $L$.

**Syntax:**
chains.pl [-v] "ca(<languages>)" < <in_file> > <out_file>

**Example:**

Query:
chains.pl -v "ca(d)" < example.chains

for getting all chains in example.chains completely edge-covered for Dutch.
**Partial Node Chains:** Giving a selection of languages, $L$, and a threshold $Min$, this query obtains the list of chains, having a subchain of length not less than $Min$ nodes, node-covered by all the languages in $L$.

**Syntax:**

```
chains.pl [-v] "pv(<languages>){<subchain_length>}" < <in_file> > <out_file>
```

where:
- `<subchain_length>`: is a natural number. It defines the minimum length of the subsequence of nodes to search.

**Example:**

Query:
```
chains.pl -v "pv(di){2}" < example.chains
```

for getting all chains in `example.chains` containing subchains of length greater or equal to 2 node-covered for Dutch and Italian.
- **Partial Edge Chains**: Giving a selection of languages, $L$, and a threshold $Min$, this query obtains the list of chains, having a subchain of length not less than $Min$ edges, edge-covered by all the languages in $L$.

**Syntax:**
```
chains.pl [-v] “pa(<languages>){<subchain_length}>” < in_file > out_file
```

**Example:**
Query:
```
chains.pl -v “pa(s){2}” < example.chains
```
for getting all chains in `example.chains` containing subchains of length greater or equal to 2 edge-covered for Spanish.

**Input:**
```
00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1000] 05636823(1000){mental_energy} [1000] 05637150(1000){libidinal_energy} [1000] 05637285(1000){catheaxis}
00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1000] 05636823(1000){mental_energy} [1000] 05636964(1100){incitement} [1100] 05637094(1100){goad}
00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100] 05633277(1111){life}
00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100] 05635834(1100){ethical_motive} [1100] 05636022(1100){hedonism}
00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100] 05635834(1100){ethical_motive} [1100] 05636133(1100){conscience} [1100] 05636402(1000){small_voice}

```
```n
```
Output:
```
00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100] 05633277(1111){life}
00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100] 05635834(1100){ethical_motive} [1100] 05636313(1100){conscience} [1100] 05636654(1100){sense_of_duty}
00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100] 056339123(1100){superego}
00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100] 05635834(1100){ethical_motive} [1100] 05636133(1100){conscience} [1100] 03839123(1100){superego}
00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1100] 05635834(1100){ethical_motive} [1000] 05636665(1000){christ_within}
00012517(1101){psychological_feature} [1100] 00013299(1111){motivation} [1000] 05634773(1000){irrational_motive} [1000] 05635682(1000){compulsion}
```
```
```n
6 subchains.
• **Partial Node Chains with Gaps**: Giving a selection of languages, \( L \), a threshold \( Min \), and a number of gaps \( G \), this query obtains the list of chains, having a subchain of length not less than \( Min \) nodes, containing \( G \) (node) gaps, node-covered by all the languages in \( L \).

**Syntax:**

\[
\text{chains.pl} [-v] \text{"pv(<languages>)(<subchain_length>)(<gaps_length>)" <in_file} > \text{<out_file>}
\]

where:

- \(<gaps_length>: \) is the number of nodes involved in the gaps of the subchain.

**Example:**

**Query:**

\[\text{chains.pl} -v \text{"pv(s){3}" < example.chains} \]

for getting all chains in \( \text{example.chains} \) containing subchains of length greater or equal to 3, with one gap, node-covered for Spanish.

**Input:**

```
00012517(1101){psychological_feature} 00013299(1111){motivation} 05636823(1000){mental_energy} 05636964(1100){incitement} 05637094(1100){goad}
00012517(1101){psychological_feature} 00013299(1111){motivation} 05636823(1000){mental_energy} 05636964(1100){incitement} 05637094(1100){goad}
00012517(1101){psychological_feature} 00013299(1111){motivation} 05636823(1000){mental_energy} 05636964(1100){incitement} 05637094(1100){goad}
00012517(1101){psychological_feature} 00013299(1111){motivation} 05636823(1000){mental_energy} 05636964(1100){incitement} 05637094(1100){goad}
00012517(1101){psychological_feature} 00013299(1111){motivation} 05636823(1000){mental_energy} 05636964(1100){incitement} 05637094(1100){goad}
```

**Output:**

```
00012517(1101){psychological_feature} 00013299(1111){motivation} 05636823(1000){mental_energy} 05636964(1100){incitement} 05637094(1100){goad}
00012517(1101){psychological_feature} 00013299(1111){motivation} 05636823(1000){mental_energy} 05636964(1100){incitement} 05637094(1100){goad}
2 subchains.
```

• **Partial Edge Chains with Gaps**: Giving a selection of languages, \( L \), a threshold \( Min \), and a number of gaps \( G \), this query obtains the list of chains, having a subchain of length not less than \( Min \) edges, containing \( G \) (edge) gaps, edge-covered by all the languages in \( L \).

**Syntax:**

\[
\text{chains.pl} [-v] \text{"pa(<languages>)(<subchain_length>)(<gaps_length>)" <in_file} > \text{<out_file>}
\]
Example:
Query:
chains.pl -v “pa(d){3}[1]” < example.chains
for getting all chains in example.chains containing subchains of length greater or equal to 3, with one gap, edge-covered for Dutch.

Input:
02657448(0111){instrument} [0001] 02010561(0101){mechanism} [0001]
02473560(0101){engine} [0001] 01991412(0111){conveyance} [0011]
03235595(0111){craft} [0001] 02051671(0111){aircraft} [0001]
02061345(0001){amphibian}
02657448(0111){instrument} [0001] 02010561(0101){mechanism} [0001]
02473560(0101){engine} [0001] 01991412(0111){conveyance} [0011]
03235595(0111){craft} [0001] 02051671(0111){aircraft} [0111]
02054514(0111){aeroplane}
02657448(0111){instrument} [0001] 02010561(0101){mechanism} [0001]
02473560(0101){engine} [0001] 01991412(0111){conveyance} [0011]
03235595(0111){craft} [0001] 02051671(0111){aircraft} [0001]
02595197(0001){hang_glider}
02657448(0111){instrument} [0001] 02010561(0101){mechanism} [0001]
02473560(0101){engine} [0001] 01991412(0111){conveyance} [0011]
03235595(0111){craft} [0001] 02051671(0111){aircraft} [0111]
02106213(0101){airship}
Output:
02657448(0111) {instrument} [0001] 02010561(0101){mechanism} [0001]
02473560(0101){engine} [0001] 01991412(0111){conveyance} [0011]
03235595(0111){craft} [0001] 02051671(0111){aircraft} [0001]
02054514(0111){aeroplane}
1 subchains.

- ‘chains2pairs.awk’: this program gets the complete chains and writes all its edges.

Syntax:
chains2pairs.awk <in_file> > <out_file>

Example:
Query:
chains2pairs.awk example.chains > example.pairs
Input:
00002403 00004262
00002403 00004022 00682831
00002403 00005489 00004885
00002403 00004885
Output:
00002403 00004262
00002403 00004022 00682831
00004022 00005489
00004022 00682831

- ‘cicles.pl’: this program takes as input a list of pairs and shows all the cycles of the list.

Syntax:
cicles.pl <in_file> > <out_file>

Example:
Query:
cicles.pl example.pairs
Input:
00002403 00002909
00002403 00005260
01121367 01046072
01046072 01121367
01121367 01121367
Output:
01121367 01046072 01121367

• `graf.pl`: this program projects a set of wordnets over another and writes the projection to a file in order to perform posterior queries. The chains generated have the next line format:

\[
\text{ili_record}((\text{english_ili}?)\langle\text{spanish_ili}?>\langle\text{dutch_ili}?>\langle\text{italian_ili}?>)
\]

\[
[\text{english_edge}?>\langle\text{spanish_edge}?>\langle\text{dutch_edge}?>\langle\text{italian_edge}?>]
\]

\[
\text{ili_record}((\text{english_ili}?)\langle\text{spanish_ili}?>\langle\text{dutch_ili}?>\langle\text{italian_ili}?>)
\]

... 

Syntax:
`graf.pl <base_wn> [projected_wn ...]`

where:

- `<base_wn>`: is the first letter of the language of the wordnet skeleton we want to load. This letter can be: `e` for English, `s` for Spanish, `d` for Dutch and `i` for Italian.
- `<projected_wn>`: is the first letter of the language of the wordnet we want to project. This letter can be: `e` for English, `s` for Spanish, `d` for Dutch and `i` for Italian.

Example:
Query:
`graf.pl d`

Input: (we suppose that dutch wordnet is only the next pairs for this example).
00002403 00002909
00002403 00005260
00005260 00513550
00005260 00739927
00005260 01219174
Output:
00002403(0010) [0010]
00002909(0010)
00002403(0010) [0010]
00005260(0010) [0010]
00513550(0010)
00739927(0010)
01219174(0010)

• `inclusions.awk`: this program deletes all the chains not finishing in a leaf.

Syntax:
`inclusions.awk <in_file> > <out_file>`

Example:
Query:
inclusions.awk example.chains > cp.chains

Input:
0013338 01472320
0013338 01472320 01257491
0013338 01472320 01257491 00086015
0013338 01472320 01257491
0013338 01472320 01257491 00202465
Output:
0013338 01472320 01257491 00086015
0013338 01472320 01257491 00202465