
EUROWORDNET: A MULTILINGUAL DATABASE OF AUTONOMOUS AND LANGUAGE-SPECIFIC WORDNETS CONNECTED VIA AN INTER-LINGUAL-INDEX.

Piek Vossen

Abstract

This paper describes the multilingual design of the EuroWordNet database. The EuroWordNet database stores wordnets as autonomous language-specific structures that are interconnected via an Inter-Lingual-Index (ILI). In this paper, we discuss the possibilities to create mappings from each wordnet to the central ILI and how the ILI itself can be adapted to provide more overlap across the wordnets. We will argue that the ILI can be condensed to a more universal index of meaning, while the wordnets can still encode any fine-grained lexicalizations for each language.

1 Introduction

EuroWordNet¹ (Vossen 1998) was a 3-year project that developed a multilingual database with wordnets for 8 European languages: English, Dutch, Italian, Spanish, French, German, Czech and Estonian. Each wordnet is structured along the same lines as the Princeton WordNet (Fellbaum 1998). WordNet contains information about nouns, verbs, adjectives and adverbs in English and is organized around the notion of a synset. A synset is a set of words with the same part-of-speech that can be interchanged in a certain context. For example, {violin; fiddle} form a synset because they can be used to refer to the same concept, but {violin; violist; fiddler} represent another

concept. It is thus clear that the same word can refer to multiple concepts (polysemy) and multiple words can point to the same concept (synonymy). Finally, synsets are related to each other by semantic relations, such as hyponymy (type-of relation between specific and more general concepts), meronymy (part-of relation between parts and wholes), etc.

The wordnets in EuroWordNet are considered as autonomous language-specific ontologies. Each language has its own set of concepts based on the lexicalisation in that language. In addition, the wordnets are interconnected via a so-called Inter-Lingual-Index so that you can go from a synset in one language to the synsets in any of the other languages. The purpose of the Inter-Lingual-Index (ILI) is to provide an efficient mapping across the wordnet structures. Since each wordnet is a separate ontology, the ILI itself can be reduced to a condensed and universal index of meaning. We will argue that such an index is better to relate wordnets to each other and to allow for the distributed development of wordnets in the world.

In the next section, we will first argue why the wordnets need to be autonomous language-specific structures. Section 3 then gives an overview of the design of the database that makes it possible to store these structures and to connect them via the ILI. The different equivalence relations that can be expressed to the ILI are explained in section 4. Finally, we will discuss the possible way in which the ILI can be adapted to provide a mapping across the wordnets.

2. Wordnets as autonomous conceptual structures

In its purest form, a wordnet is a network of concepts based on the words from a language. This means that the fund of concepts is based on the vocabulary. There are no concepts for which there are not words or expressions in a language. Another typical feature of a wordnet is that words are mainly defined by the relations to each other. Via the relations, an ontological structure is

expressed about beliefs and properties of the world, as encapsulated in the vocabulary of the language.

The EuroWordNet database contains many wordnets that are interconnected. The multilinguality of the EuroWordNet database raises a fundamental issue with respect to the status of the monolingual structures of each wordnet. What should be done with differences in the ontological structure across the wordnets? If 'equivalent' words are related in different ways in the different resources, we have to make a decision about the legitimacy of these differences. For example, figure-1 represents two fragments of wordnets for Dutch and English, where the concepts and their relations are very different. The Dutch conceptual structure seems to be much simpler than the English structure. There are two main reasons for this:

- (1) The concepts 'container', 'artifact object', 'instrumentality' and 'natural object' are not lexicalised in Dutch, thus leading to less hierarchical levels;
- (2) The concept 'instrumentality' and 'natural object' seem to be added in WordNet 1.5 to group certain other concepts and are not motivated by the vocabulary of the English language.

INSERT FIGURE 1

Apparently, the ontological structure of wordnets can be fundamentally different across languages. This is a necessary consequence of the different lexicalisation patterns. This does not mean that the concept of 'container' is unknown in the Dutch culture. It only means that it is not lexicalised. Dutch people cannot refer to 'container' concepts in general using a single word.² The EuroWordNet database contains the vocabularies of 8 languages, while more and more wordnets are being developed for many more languages all over the world. The Global Wordnet Association (<http://www.globalwordnet.org>) so far has registered wordnets for more than 35

different languages. It is thus no option to try to unify all these different ontological structures and likewise build a database where all languages share the same ontology.

Instead, EuroWordNet took the decision to store the wordnets as autonomous ontological structures. Each wordnet in a language is a unique lexicalisation pattern that represents a certain structuring of the world. Preserving the implicit ontological structure of each language is a unique feature of the database. The wordnets are seen as linguistic ontologies rather than ontologies for making inferences only. They are 'wordnets' in the true sense of the word and therefore capture valuable information about conceptualisations that are lexicalised in a language: what is the available fund of words and expressions in a language, and what words and expressions can substitute each other (Cruse 1986).

3. The design of the EuroWordNet database

To maintain the language-specific structures, the EuroWordNet database makes a distinction between the language-specific modules and language-independent modules, as schematically shown in figure-2 below. Each language module represents an autonomous and unique language-specific system of language-internal relations between synsets. It consists of its own fund of concepts with relations and a lexical items table that points to these concepts. Equivalence relations across these synsets are established via a so-called Inter-Lingual-Index. The Inter-Lingual-Index is a fund of concepts with the sole purpose of connecting concepts across wordnets. Most concepts in each wordnet are related to the closest concepts in that Inter-Lingual-Index. Concepts across languages that are related to the same index concepts can be considered as alternative vocabulary items. This is illustrated for the language-specific synsets linked to the ILI-record drive. Via the Inter-Lingual-Index (ILI), it is also possible to share other knowledge that is language neutral. In figure-2, we see a domain hierarchy (Magnini and Cavigliá 2000) and a formal semantic ontology

(Guarino 1998, Niles and Pease 2003) that is associated with the concept index. Such structures can be used by any language that is related to the ILI.

The ILI is an unstructured list of meanings, mainly taken from WordNet1.5, where each ILI-record consists of a synset, an English gloss specifying the meaning and a reference to its source. The only purpose of the ILI is to mediate between the synsets of the language-specific wordnets. No relations are therefore maintained between the ILI-records as such. As an unstructured list, there is no need to discuss changes or updates to the index from a many-to-many perspective. Note that it will nevertheless be possible to indirectly see a structuring of a set of ILI-records by viewing the language-internal relations of the language-specific concepts that are related to the set of ILI-records. Since WordNet1.5 is linked to the index in the same way as any of the other wordnets, it is still possible to recover the original internal organization of the synsets in terms of the semantic relations in WordNet1.5. In fact, any wordnet or ontology linked to the ILI can be seen as a possible ontological structuring of the associated concepts.

The EuroWordNet model thus subsumes the SIMULLDA model described by Janssen (this volume) but is more powerful. EuroWordNet allows for ANY structuring of the set of concepts (ILI) that is used to interlink the meanings across languages. Likewise, EuroWordNet does not impose a single structuring and it does not require full consensus for the expressed structures across all the languages. The SIMULLDA database imposes a single structure on all languages and thus the same substitution relations between the words of languages that follow from such structures. The latter is fundamentally wrong.

The advantages of an interlingua such as the Inter-Lingua-Index are well-known in machine translation systems (Nirenburg 1989, Copeland et al. 1991):

- (1) It is not necessary to specify many-to-many equivalence relations between each language-pair and to have consensus across all the groups on the equivalence relations: each group only considers the equivalence relations to the Index.
- (2) New languages can be added without having to reconsider the equivalence relations for the other languages.
- (3) It is possible to adapt the Inter-Lingual-Index as a central resource to make the matching more efficient or precise, even without consulting the builders of the wordnets in different languages.

Points 1 and 2 have been demonstrated by the BalkaNet project (www.ceid.upatras.gr/Balkanet) that independently extended the database with wordnets for six other languages, without consultation of the former group. In this paper, we will discuss the last point. As we will see below, the unstructured and central ILI makes it possible to reconsider the equivalence relations rather easily.

4 Equivalence relations in EuroWordNet

The equivalence relations between synsets in each language and the Inter-Lingual-Index are to a large extent parallel to the language internal relations that are expressed between concepts in each wordnet. The most important relation is EQ_SYNONYM, which only holds if there is a 1-to-1 mapping between synsets and ILI-records. In addition, there are relations for complex-equivalence relations, among which the most important are:

EQ_NEAR_SYNONYM when a meaning matches multiple ILI-records simultaneously, when multiple synsets match with the same ILI-record, or when there is some doubt about the precise mapping.

EQ_HAS_HYPERONYM when a meaning is more specific than any available ILI-record.

EQ_HAS_HYPONYM when a meaning can only be linked to more specific ILI-records.

In EuroWordNet, the complex relations are needed when there is a lexical gap in one language or when meanings do not exactly fit.³ The first situation, in which a single synset matches several ILI-records simultaneously, occurs quite often. For example, in the Dutch resource there is only one sense for *schoonmaken* (to clean) which simultaneously matches with at least 4 senses of clean in the ILI (based on WordNet1.5):

- {make clean by removing dirt, filth, or unwanted substances from}
- {remove unwanted substances from, such as feathers or pits, as of chickens or fruit}
- {remove in making clean; 'Clean the spots off the rug'}
- {remove unwanted substances from - (as in chemistry)}

The Dutch synset *schoonmaken* will thus be linked with an EQ_NEAR_SYNONYM relation to all these senses of clean. The reverse situation also occurs. For example, *versiersel* and *versiering* are not coded as synonyms in the Dutch resource but they can still both be linked to the same ILI synset decoration. They share the same ILI-record but the equivalence relation should be EQ_NEAR_SYNONYM to indicate a mismatch.

The EQ_HAS_HYPERONYM is typically used for gaps in WordNet1.5 or in English. Such gaps can be cultural or pragmatic. A cultural gap is a concept not known in the English/American culture, e.g. the Dutch noun *citroenjenever*, which is a kind of gin made out of lemon skin, or the Dutch verb: *klunen* ('to walk on skates over land'). Pragmatic gaps are caused by lexicalization differences between languages, in the sense that in this case the concept is known but not expressed by a single lexicalized form in English, e.g.:

Dutch: *doodschoppen* ('to kick to death'),

Spanish: *alevín* ('young fish'),

Italian: *rincasare* ('to go back home').

In these cases the lexicalization patterns in the languages are different from English but the concepts are familiar to all cultures. Typically, a concept like *doodschoppen* ('kick to death') in Dutch will get two eq_hyperonym relations, one to 'to kill' and one to 'to kick'. This is parallel to the multiple hyperonyms the word will receive in the Dutch wordnet as language-internal relations. Similarly, Spanish 'alevín' (young fish) can both be linked with an eq_hyperonym to 'fish' and eq_be_in_state to 'young'. Using multiple equivalence relations the meanings of many synsets can be exhaustively linked to the ILI.

In all the above cases, the non-English word is more specific and thus can be related to a more general English ILI-concept using at least an EQ_HAS_HYPERONYM relation and possibly other relations. The EQ_HAS_HYPONYM is then used for the reversed situation, when WordNet1.5 only provides more narrow terms. An example is Spanish *dedo* which can be used to refer to both finger and toe.

5. Towards a universal index of meaning

Equivalence relations can be rather complex in the EuroWordNet database. As the above examples make clear, there are many subtle differences between the languages and also between the resources that have been used to build the wordnets. Some of these differences are arbitrary and some of them are linguistically motivated and should be taken seriously. Nevertheless, the design of the EuroWordNet database makes it possible to reconsider the equivalence relations from a central point of view. It is possible to adapt and modify the ILI just to make the mapping of the wordnets more efficient. There are two opposing approaches that can be considered:

- (1) Maximize the number of concepts so that the ILI is always the superset of the concepts occurring in all the other wordnets.
- (2) Minimize the number of concepts to a set of essential concepts that are sufficient to establish equivalence relations across synsets.

The first approach requires that developers of wordnets follow a strict procedure for all synsets that they cannot relate with a 1:1 mapping to the ILI. They will have to make a proposal to extend the ILI with a new concept, and a central group needs to judge these proposals, verify if there are overlapping proposals and make the changes to the ILI. When a new ILI is released, all the involved groups have to reconsider the equivalence relations of concepts in their wordnet that can be affected by these changes. It may be clear that this is a time-consuming and complex procedure that requires that the separate wordnets need to be updated continuously. This approach is suggested by (Janssen, this volume) for the SIMULLDA database. A structured ILI requires consensus across all the different languages and thus continuous consultation about restructurings.

In EuroWordNet, we investigated the changes to the ILI that would be involved in this approach (Vossen et al. 1999). We inspected a sample of the Italian and German mismatches to see if they could potentially overlap with Dutch synsets. A random sample of 36 German noun synsets showed that 50% of the nouns (18) have an equivalent in Dutch. For a sample of 59 Italian noun synsets there is at least an overlap of 30% (20) with Dutch. Examples are: *Arbeitszeitverkürzung* (DE) = *arbeidstijdverkorting* (NL) = ('a reduction of working hours') and *baita* (IT) = *berghut* (NL) = ('a cabin in the mountain'). If we roughly quantify these results for the total Dutch wordnet, where about 6,000 Dutch noun synsets could not be linked to the ILI, this would imply that at least 30% (2,000 synsets) represent new concepts that overlap with German or Italian, and therefore could be added to the ILI.

However, the inspection of the mismatches also revealed that most of the ‘new’ concepts are rather specific and their meanings are often predictable. Especially, in the case of systematic differences it hardly makes sense to extend the ILI with these concepts. For example, German and Dutch productively lexicalize verb compounds that imply both the way in which a change comes about and the resulting change. Consider the following Dutch examples which can easily be extended with hundreds of verbs:

- *droogmaken* (literally dry make, ‘to dry’), *droogwrijven* (‘to dry by rubbing’), *droogvegen* (‘to dry by sweeping’), etc.
- *fijnmaken* (literally fine make, ‘to crush’), *fijnwrijven* (‘to crush by rubbing’), *fijnstampen* (‘to crush by stamping’), etc.
- *doden* (‘to kill’), *doodschoppen* (‘to kick to death’), *doodslaan* (‘to beat to death’), *doodstampen* (‘to stamp to death’), etc.

Something similar can be said for gender differences, which are systematically lexicalized in Roman languages and partially lexicalized in Germanic languages but are not differentiated in English, or aspectual phases, which are strongly lexicalized in Slavic languages and not in English..

The maximalist approach would require a systematic extension of the ILI for all these cases. The minimalistic approach is instead more conservative to change the ILI. The ILI is only extended for rather specific non-productive and non-predictable concepts, such as *Arbeitszeitverkürzung* (‘reduction of working hours’). Most of the differences of languages are however rather systematic and predictable. In these cases, it will be sufficient to relate the concepts to multiple ILI-records with multiple complex equivalence relations that exhaustively define the concept. As figure-3 shows, the ILI itself can remain rather condensed and abstract and can still be used to uniquely relate concepts in languages and even express equivalence relations cross languages.

INSERT FIGURE 3

The Dutch verb compounds in figure 3 can exhaustively be related to multiple concepts in the ILI. If there is a similar unique mapping from another language to the same concepts (as is the case for German here), a unique equivalence relation can still be derived even if the concept itself is not uniquely present in the ILI. The same holds for female variants such as Spanish ‘cajera’ and Dutch ‘caissière’ (‘cash girl’) and any other cases of lexical incorporation such as Spanish ‘alevín’ (young fish). Lexical incorporation is a powerful and productive system in all languages. It yields massive differences in concepts and meanings.

The minimalistic approach also opens a completely different perspective on the ILI. The ILI itself can be further reduced by abstracting from various sense-distinctions that are now based on the lexicalisation in English. Some of these sense-distinctions are arbitrary, as we have seen for the different meanings of ‘clean’, and can be easily generalized. Other distinctions follow from systematic lexicalisation patterns in English. As many scholars have suggested (Apresjan 1973; Copestake and Briscoe 1991; Nunberg and Zaenen 1992, Levin 1993), many sense distinctions in English are regular, either based on metonymic relations (embassy as a building and an institute) or on diathesis alternations (to open a door with a key versus the key opened the door). This can even be said for many derivational concepts in the ILI. Strictly speaking there is no need to have separate concepts for ‘depart’ as a verb and ‘departure’ as a noun.

EuroWordNet so far followed the minimalistic approach. Rather than extending the ILI, we tried to globalise the concepts in the ILI to enlarge the mapping of synsets across languages and to make it easier to link synsets to the ILI. Globalisation of the ILI can be done rather easily with so-called Composite ILI-records, which can be compared with Complex Types as defined by Pustejovsky (1995). These records group more specific existing ILI records. The grouping can be differentiated as a metonymic relation, a generalization or a diathesis alternation. The composite ILI-records can be derived automatically from the sense-patterns in English and the structural

properties of WordNet1.5 (Buitelaar 1998, Peters et al. 1998). They can also be derived from the equivalence relations expressed by the other wordnets. Whenever synsets are related to multiple ILI-records, the ILI-records can also be grouped together by a composite ILI-record. This is shown in the next example.

INSERT FIGURE 4

Here we see that the Dutch equivalent for ‘university’ is related to the institute meaning, whereas the Spanish equivalent is linked to university as a building. By adding a Composite ILI that groups both ‘university’ concepts by metonymy, we can derive a metonymic mapping across Dutch and Spanish indirectly from the fact that these words are related to an element from a metonymic group.

The composite ILI-records can be added to the database independently of the mappings of the other wordnets. By adding the composite ILI-records, we more or less build an index on top of an index. The existing equivalence relations can then still be considered but there is also a possibility to match synsets across languages on a more global index level, represented by the composite ILI records. This has been done for certain cases in EuroWordNet, which increased the mapping across languages with 5% for nouns (Vossen et al 1999).

It is possible in the future to apply this process more drastically and derive a smaller set of ILI-records that is more condensed and universal. This is suggested in figure 4.

INSERT FIGURE 5

Starting from the 90.000 concepts from WordNet1.5 in the ILI, we could gradually reduce the concepts by generalizing over the metonymic variants, the diathesis alternations and the meaning specializations. Furthermore, we can exclude rather rich domain specific extensions that are currently in WordNet, e.g. the biological classes of animals and plants. Next, we can exclude all derivations with predictable meanings and make the concepts in the ILI part-of-speech neutral so that other languages can map to equivalences regardless of the part-of-speech.

Such a condensed index will make it much easier for other wordnet developers to link their resources. Natural Language Applications will also benefit from the more global differentiation of concepts in the ILI (Vossen et al. 1999), since many systems report difficulties to use the current fine-grained distinctions.

6. Accessing equivalence relations through the ILI

An important characteristic of the equivalence relations is that they are established at the synset level. This is different from a traditional bilingual dictionary where specific relations are expressed between individual words or word-senses. EuroWordNet thus matches concepts rather than words and multiple concepts may share ILI-records (index-terms) or single concepts may yield multiple ILI-records. The database thus provides the possibility to project a single concept or a cluster of concepts to another language, either specifically or in a more fuzzy way. Table 1 gives an overview of these mappings. Once we have accessed a cluster of concepts in the target language, we can further use the language-internal relations to see the conceptual dependencies between these words (and possibly other words).

INSERT TABLE 1

Compared to bilingual dictionaries, the EuroWordNet database gives a more-comprehensive overview of concept-lexicalisation in the target language, from which to choose the best candidate. In this sense, we can make a parallel with the 'Shake and Bake' methodology in Machine Translation (Whitelock 1992), where first an abstraction is made from the structural properties in the source language to a more neutral conceptual level (shake), and next a (possibly different) new structure is generated in the target language (bake). In the case of EuroWordNet, we are dealing with lexical shake: abstract from the lexicalization that may be specific for a language. Bake is then

possible by selecting the most appropriate candidate on the basis of co-occurrence restrictions in the target language, or the pragmatic and morpho-syntactic properties of the members in the synset.

7. Conclusions

We described the design of the EuroWordNet database. We argued that it is necessary that wordnets are represented as autonomous language-specific ontologies. For creating equivalence relations, the databases use an unstructured Inter-Lingual-Index (ILI). The purpose of this index is to provide an efficient mapping across the autonomous wordnets. We explained that such an indexed can be rather condensed and universal. It is not

necessary to include concepts for any lexicalised word in any language. Especially for the future, when many more wordnets are added to the databases, such a condensed and universal index will be beneficiary. Finally, we explained that the fuzzy and more global equivalence mappings can still be used to derive correct translation equivalences across languages, using a kind of lexical shake-and-bake method.

Notes

1. EuroWordNet (LE2-4003 and LE-8328) was funded by the European Community within the Telematics Application Programme of the 4th Framework (DG-XIII, Luxembourg). The project started March 1996 and ended July 1999.
2. The EuroWordNet model is a weaker version of the Sapir-Whorf hypothesis of linguistic relativity (Whorf, 1956). Whereas Sapir and Whorf claimed that people think differently because of the differences across languages, EuroWordNet claims at most that people talk differently.
3. Compare the kinds of mismatches across word meanings captured in the Acquilex project by complex TLINKS (Ageno et al 1993, Copestake et al. 1995).

4. More complicated cases, where the meaning distinctions partially match with the ILI-records and where people encoded fuzzy EQ_NEAR_SYNONYM relations, are not considered here.

References

- Ageno A., F. Ribas, G. Rigau, H. Rodriguez and F. Verdejo.** 1993. 'TGE: Tlinks Generation Environment'. *Acquilex II (BRA 7315) Working Paper 7*. Polytechnica de Catalunya, Barcelona.
- Apresjan, J.** 1973. 'Regular Polysemy'. *Linguistics* 142: 5-32.
- Buitelaar, P.** 1998. *Corelex: Systematic Polysemy and Underspecification*. PhD., Department of Computer Science, Brandeis University.
- Copeland, C., Durand, J., Krauwer, S. and Maegaard, B. (eds.)**. 1991. *The Eurotra Formal Specifications*,. Office for Official Publications of the European Community, Luxembourg.
- Copestake A. and Briscoe, T.** 1991. 'Lexical operations in a unification-based framework'. In Pustejovsky J. and Bergler S. (eds), *Lexical Semantics and Knowledge Representation*, First SIGLEX Workshop, Berkeley, CA, Springer-Verlag, 101 - 118
- Copestake A., T Briscoe, P. Vossen, A Ageno, I Castellon, F Ribas, G Rigau, H Rodriguez, A Sanmiotou.** 1995. 'Acquisition of Lexical Translation Relations from MRDs'. *Journal of Machine Translation* 9 (3): 183-219.
- Cruse, D. A.** 1986. *Lexical Semantics*. Cambridge, Cambridge University Press.
- Fellbaum, C. (ed.)**. 1998. *WordNet. An Electronic Lexical Database*. The MIT Press 1998.
- Guarino, N.** 1998. 'Some Ontological Principles for Designing Upper Level Lexical Resources' In A. Rubio, N. Gallardo, R. Catro and A. Tejada, *Proceedings of First International Conference on Language Resources and Evaluation*, Granada, 527-534.
- Levin, B.** 1993. *English Verb Classes and Alternations, a Preliminary Investigation*. University of Chicago Press, Chicago and London.

- Magnini, B. and G Cavagliá.** 2000. 'Integrating subject field codes into wordnet.' In *Proceedings of the Second International Conference on Language Resources and Evaluation LREC'2000*, Athens, 1413-1420.
- Niles, I. and A. Pease.** 2003. 'Mapping WordNet to the SUMO ontology'. *Proceedings of the IEEE International Knowledge Engineering conference*, Las Vegas, NV, June 23-26.
- Nirenburg, S. (ed.).** 1989. 'Knowledge-based Machine Translation'. *Machine Translation* 4 (1 and 2, Special issue) Kluwer Publishers, Dordrecht.
- Nunberg, G & A. Zaenen.** 1992. 'Systematic Polysemy in Lexicology and Lexicography'. In H. Tommola, K. Varantola et al. (eds), *EURALEX'92 Proceedings*, University of Tampere, Part II, 387 – 396.
- Peters, W., I. Peters, and P. Vossen.** 1998. 'The Reduction of Semantic Ambiguity in Linguistic Resources'. In A. Rubio, N. Gallardo, R. Catro and A. Tejada (eds), *Proceedings of First International Conference on Language Resources and Evaluation*, Granada, 409-416
- Pustejovsky, J.** 1995. *The Generative Lexicon*, MIT Press, Cambridge MA.
- Vossen, P. (ed).**1998. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks*, Kluwer Academic Publishers, Dordrecht.
- Vossen, P. Peters, W., J. Gonzalo.** 1999. 'Towards a Universal Index of Meaning'. *Proceedings of the ACL-99 Siglex workshop*, University of Maryland, 81-90
- Whitelock, P.** 1992. 'Shake-and-bake translation', *Proceedings of the 14th International Conference on Computational Linguistics (COLING '92)*, Nantes, 23-28 August 1992', Vol. 2, University of Nantes, Nantes, 784-790.
- Whorf, B.L.** 1956. 'Science and linguistics' In J.B. Carroll (ed), *Language, thought and reality: Selected writings of Benjamin Lee Whorf*. Cambridge Mass., MIT Press, 1956, 207-219.

Figure 1: Different conceptual structures in wordnets

Figure 2: The multilingual architecture of the EuroWordNet database

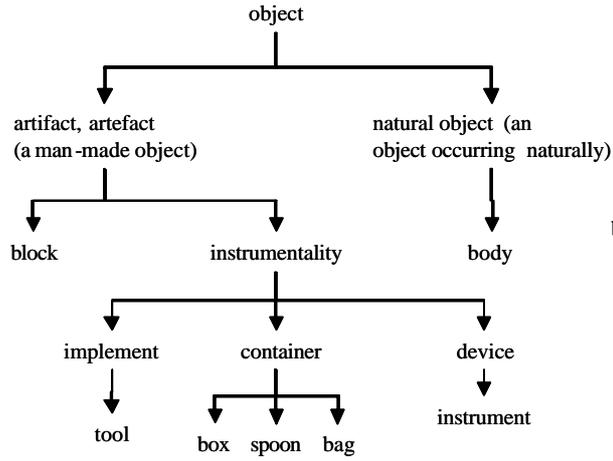
Figure 3: Productive and compositional concepts exhaustively linked

Figure 4: : Increased mapping via Complex ILI records

Figure 5: Towards a universal index of meaning

Table 1 : Overview of mapping relations to the ILI

English Wordnet 1.5



Dutch Wordnet

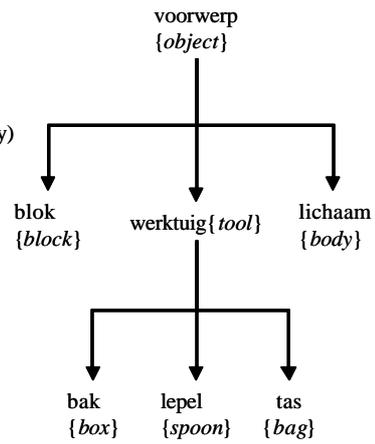


Figure 1: Different conceptual structures in wordnets

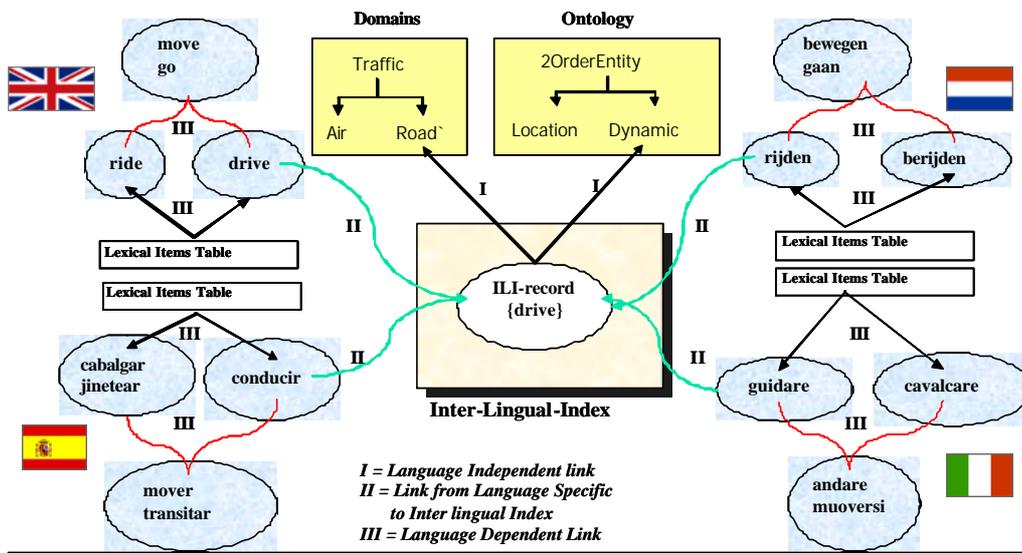


Figure 2: The multilingual architecture of the EuroWordNet database

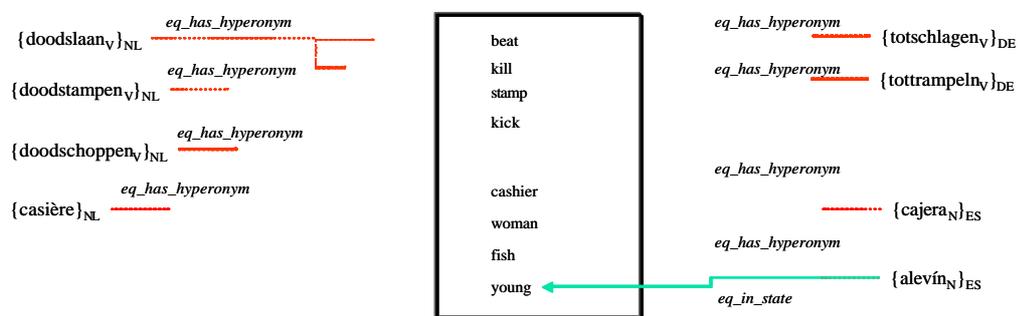


Figure 3: Productive and compositional concepts exhaustively linked

Formatted: English (U.K.)

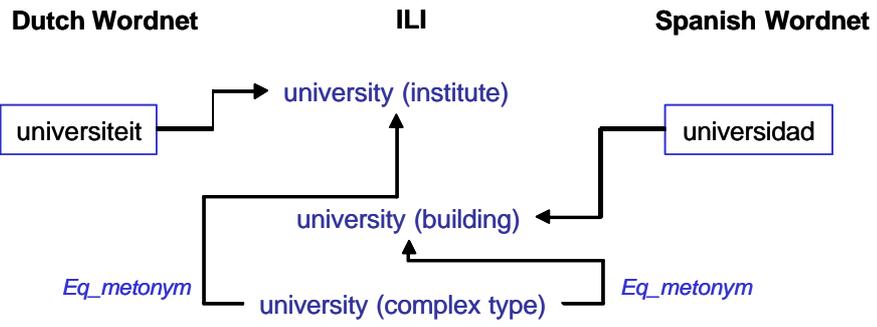


Figure 4: Increased mapping via Complex ILI records

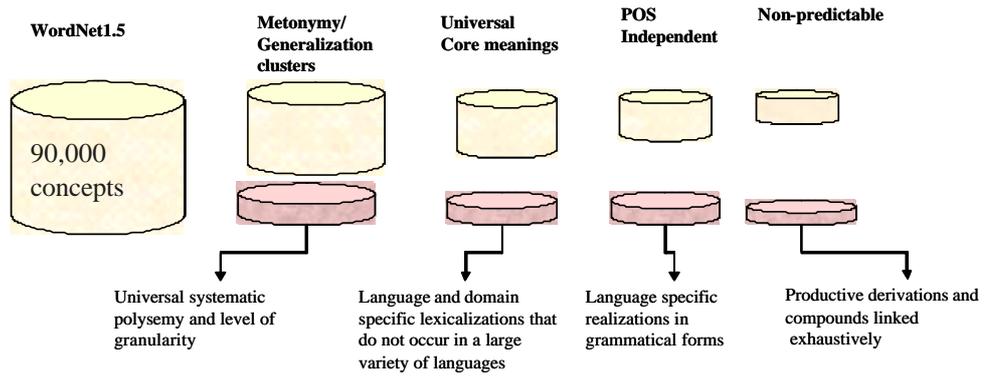


Figure 5: Towards a universal index of meaning

Formatted: English (U.K.)

Table 1 : Overview of mapping relations to the ILI

<i>Relation</i>	<i>POS</i>	<i>Source: Target</i>	<i>Example</i>
eq_synonym	same	1:1	auto : car
eq_near_synonym	any	many : many	apparaat, machine, toestel : apparatus, machine, device
eq_hyperonym	same	many : 1 (usually)	citroenjenever: gin
eq_hyponym	same	(usually) 1 : many	dedo : toe, finger
eq_metonymy	same	many/1 : 1	universiteit, universiteitsgebouw: university
eq_diathesis	same	many/1 : 1	raken (cause to hit), raken:hit
eq_generalization	same	many/1 : 1	schoonmaken : clean