

# EuroWordNet

Funded by the European Commission, Directorat General XIII, Luxembourg

Application Area:	<b>Language Resources, Language Engineering</b>
Start Date:	<b>March 1996</b>
Duration:	<b>36 Months</b>
Total Effort:	<b>149 Person Months</b>
Budget:	<b>900 KECU</b>

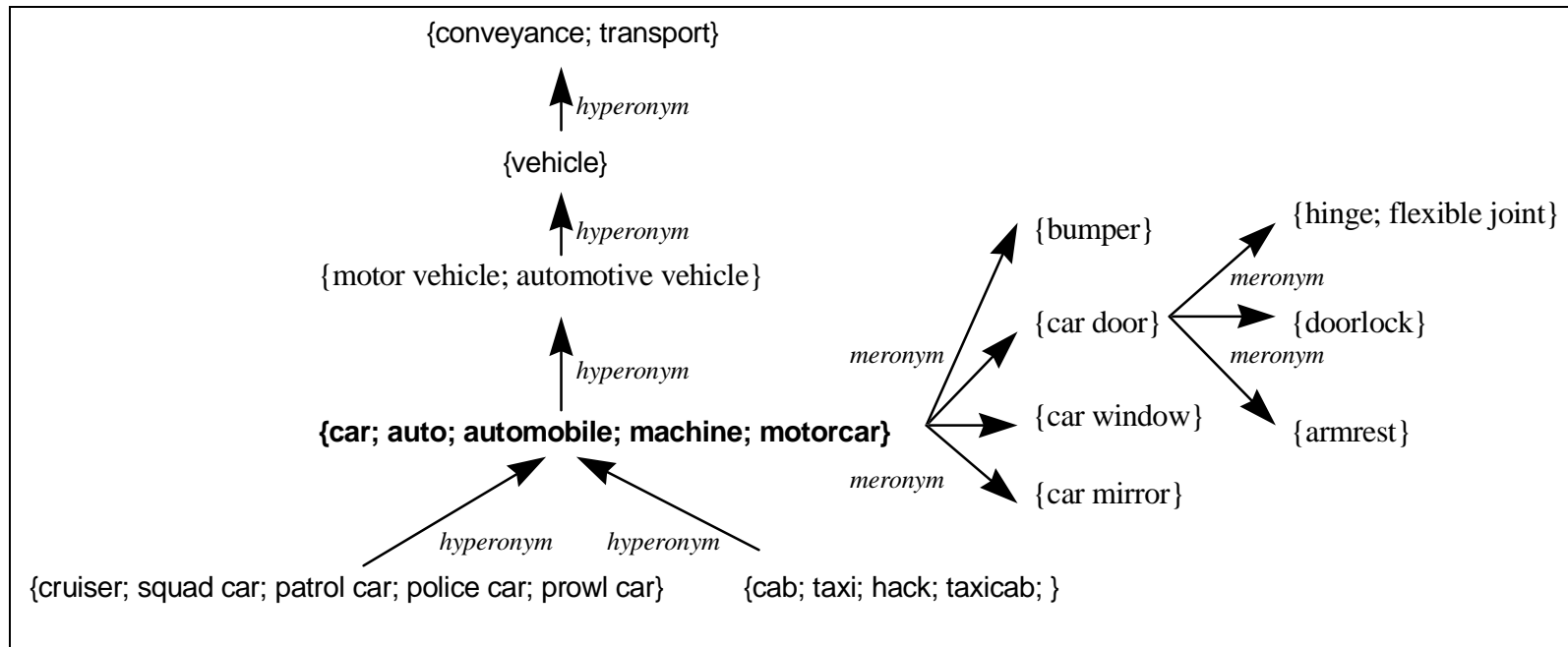
- *the development of wordnets with the basic semantic relations for several European languages.*
- *each meaning will be linked to a meaning in the Princeton WordNet.*
- *the wordnets with their equivalence relations will be loaded into a common multilingual database.*
- *size of vocabulary: 30,000 concepts, 50,000 word senses*
- *type of vocabulary:*
  - *the most frequent words of the languages*
  - *all concepts needed to relate more specific concepts*
- *freely available on the basis of users-licenses*

**Currently covered languages:** *English, Dutch, Spanish and Italian.*

**Extension 1998:** *German, Swedish, French, Czech, Estonian.*

Princeton WordNet (Miller et al 1990) is semantic network:

- organized around the notion of synsets = sets of synonymous words
- basic semantic relations between these synsets



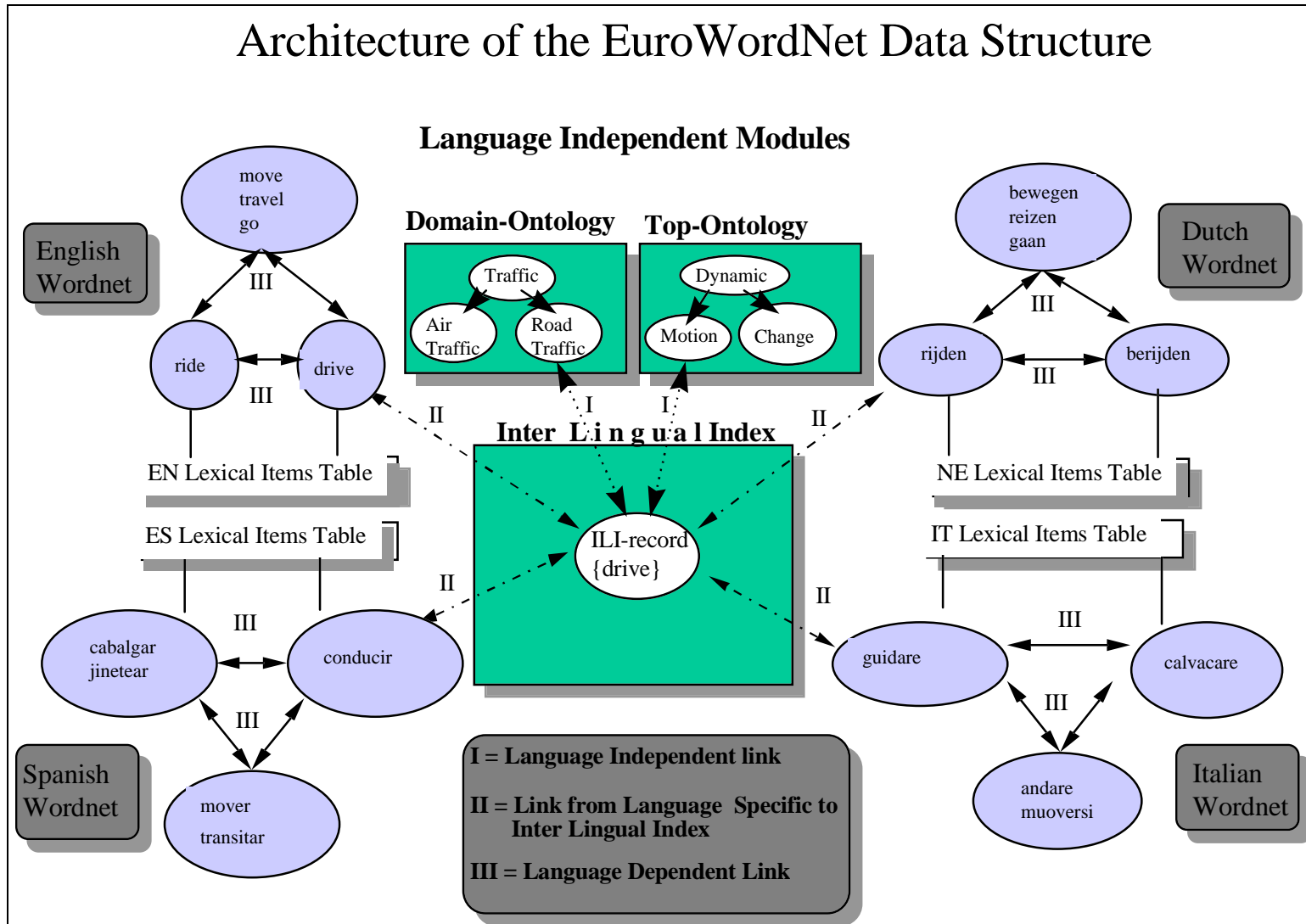
## Specific features:

- *each wordnet reflects the relations as a language-internal system, maintaining cultural and linguistic differences in the wordnets.*
- *using the multilingual relations it possible to go from one language to another language, which makes it possible to compare the wordnets to track down inconsistencies and cross-linguistic differences.*
- *each wordnet is linked to a language independent top-ontology and domain labels which can be used to customize the multilingual database without having to access the language specific wordnets.*

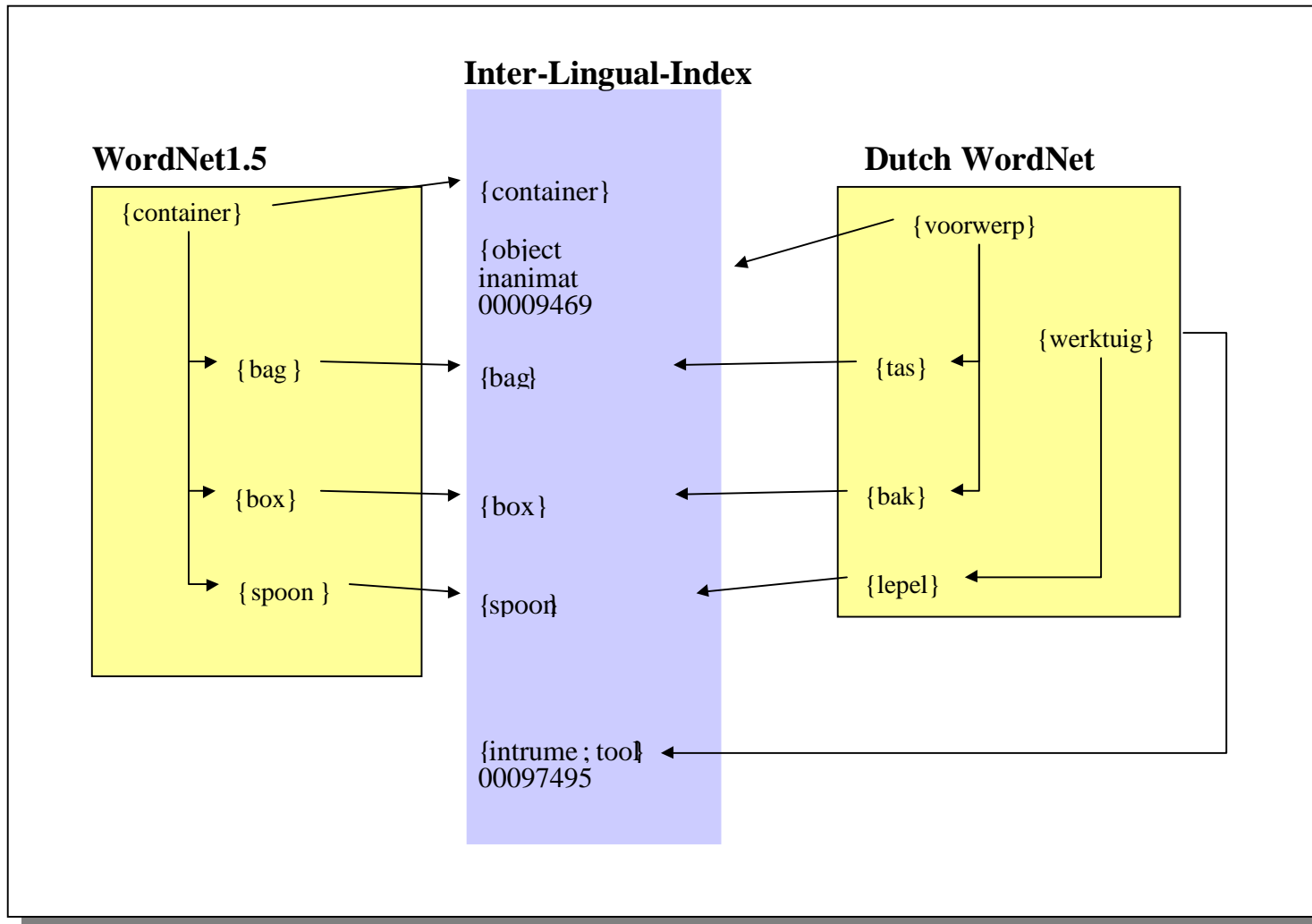




# Architecture of the EuroWordNet Data Structure



# Wordnets as autonomous language-specific structures



## **The wordnets as linguistic ontologies rather than conceptual ontologies:**

**Conceptual ontology:** *a particular level or structuring may be required to achieve a better control or performance, or a more compact and coherent structure. For this purpose it may be necessary to introduce artificial levels for concepts which are not lexicalized in a language (e.g. external body parts), or it may be necessary to neglect levels which are lexicalized but not relevant for the purpose of the ontology.*

**Linguistic ontology:** *exactly reflects the relations between the lexicalized words and expressions in a language. It therefore captures valuable information about the expressiveness of languages: what is the available fund of words and expressions in a language.*

**Each wordnet should thus be seen as an autonomous language-specific structure.**

**Practical motivation:** *the wordnets will be (as much as possible) derived from existing resources, databases and tools. This gives each of the sites a different starting point for building their local wordnet. It is therefore important that we allow for a maximum of flexibility in producing the wordnets and structures.*



## Linguistic Principles for deriving relations:

### 1. *Substitution tests (Cruse 1986):*

- 1 a. It is a fiddle therefore it is a violin.  
b. It is a violin therefore it is a fiddle.
- 2 a. It is a dog therefore it is an animal.  
b. \*It is an animal therefore it is a dog.
- 3 a. to kill (/a murder) causes to die (/ death)  
to kill (/a murder) has to die (/ death) as a consequence  
b. \*to die / death causes to kill  
\*to die / death has to kill as a consequence

### 2. *Principle of Economy (Dik 1978):*

If a word  $W_1$  (*animal*) is the hyperonym of  $W_2$  (*mammal*) and  $W_2$  is the hyperonym of  $W_3$  (*dog*) then  $W_3$  (*dog*) should not be linked to  $W_1$  (*animal*) but to  $W_2$  (*mammal*).

### 3. *Principle of Compatibility*

If a word  $W_1$  is related to  $W_2$  via relation  $R_1$ ,  $W_1$  and  $W_2$  cannot be related via relation  $R_n$ , where  $R_n$  is defined as a distinct relation from  $R_1$ .

## Conflicting Starting points

*1. There should be a maximum of flexibility:*

- *the wordnets should be able to reflect language-specific relations and patterns*
- *the wordnets should be built relatively independently because each sites has different starting points:*
  - *different tools, database and resources (Machine Readable Dictionaries)*
  - *differences in the languages*

*2. the wordnets have to be compatible in terms of coverage and relations to be useful for multilingual information retrieval and translations tools and to be able to compare the wordnets.*

## Measures to achieve maximal compatibility

- *the results are loaded into a common Multilingual Database:*
  - *consistency checks and types of incompatibility*
  - *specific comparison options to measure consistency and overlap in coverage*
- *user-guides for building wordnets in each language:*
  - *the steps to encode the relations for a word meaning.*
  - *common tests and criteria for all the relations.*
  - *overview of problems and solutions.*
- *a set of common Base-Concepts which are shared by all the sites, having:*
  - *most relations and the most-important positions in the wordnets*
  - *most meanings and badly defined*
- *Classification of the common Base Concept in terms of a Top-Ontology of 63 basic Semantic Distinctions*
- *Top-Down Approach, where first the Base Concepts and their direct context are (manually) encoded and next the wordnets are (semi-automatically) extended top-down to include more specific concepts that depend on these Base Concept.*

# Top-Ontology and Base Concepts

## Top-Ontology with 63 higher-level concepts

- *Existing Ontologies:*
  - *WordNet1.5 top-levels*
  - *Aktions-Art models (Vendler, Verkuyl)*
  - *Acquilex and Sift ontologies (EC-projects)*
  - *Qualia-structure (Pustejovsky)*
  - *Upper-Model, MikroKosmos, Cyc, Ad Hoc ANSI-Committee on ontologies*
- *The ontology was adapted to represent the variety of concepts in the set of Common Base Concepts, across the 4 language:.*
  - *homogenous Base-Concept Clusters*
  - *average size of Base Concept Cluster*
  - *apply to both nouns and verbs*
- *single hyponymy relations between Top Concepts, cross-classification of Base-Concept to Top-Concepts.*

**Set of 1024 common Base Concepts making up the core of the separate wordnets.**

## Base Concepts

### Procedure:

1. *Each site determined the set of word meanings with most relations (up to 15% of all relations) and high positions in the hierarchy.*
2. *This set was extended with all meanings used to define the first selection.*
3. *The local selection was translated to WordNet1.5 equivalences: 4 lists of WordNet1.5 synsets (between 450 – 2000 synsets per selection).*
4. *These sets of WordNet1.5 translations have been compared.*

### Concepts selected by all sites:

*30 synsets (24 nouns synsets, 6 verb synsets).*

### Explanations:

1. *The individual selections are not representative enough.*
2. *There are major differences in the way meanings are classified, which have an effect on the frequency of the relations.*
3. *The translations of the selection to WordNet1.5 synsets are not reliable*
4. *The resources cover very different vocabularies*

## Nominal Base Concepts selected by all four sites

act 1	cloth 1	flora 1	mental attitude 1	solid ground 1
activity 1	cognition 1	food 1	mortal 1	someone 1
amount of time 1	compound 4	ground 7	nutrient 1	soul 1
animal 1	construction 4	human 1	ornament 1	structure 1
animate being 1	creature 1	human action 1	period 3	stuff 7
attitude 3	decoration 2	human activity 1	period of time 1	substance 1
beast 1	drink 2	individual 1	person 1	terra firma 1
beverage 1	dry land 1	knowledge 1	phenomenon 1	textile 1
brute 1	earth 3	land 6	plant 1	time period 1
chemical	element 6	line 26	plant life 1	worker 2
compound 1	fabric 1	material 1	point 12	
chemical element 1	fauna 1	material 5	potable 1	
	feeling 1	matter 1	quality 1	

## Verbal Base Concepts selected by all four sites

be 4  
cause 6  
cover 16  
create 2  
get 9  
go 14  
have 7  
have the quality of being 1  
induce 2  
locomote 1  
make 12  
make 13  
move 15  
remove 2  
stimulate 3  
take 4  
take away 1  
travel 4

## Concepts selected by at least two sites

### Intersections of pairs

<i>Nouns</i>	<i>AMS</i>	<i>FUE</i>	<i>PSA</i>	<i>SHE</i>
<b>AMS</b>	<b>1027</b>	<b>103</b>	<b>182</b>	<b>333</b>
<b>FUE</b>	<b>103</b>	<b>523</b>	<b>45</b>	<b>284</b>
<b>PSA</b>	<b>182</b>	<b>45</b>	<b>334</b>	<b>167</b>
<b>SHE</b>	<b>333</b>	<b>284</b>	<b>167</b>	<b>1296</b>

<i>Verbs</i>	<i>AMS</i>	<i>FUE</i>	<i>PSA</i>	<i>SHE</i>
<b>AMS</b>	<b>323</b>	<b>36</b>	<b>42</b>	<b>86</b>
<b>FUE</b>	<b>36</b>	<b>128</b>	<b>18</b>	<b>43</b>
<b>PSA</b>	<b>42</b>	<b>18</b>	<b>104</b>	<b>39</b>
<b>SHE</b>	<b>86</b>	<b>43</b>	<b>39</b>	<b>236</b>

### Total Set of shared Base Concepts

	<i>Nouns</i>	<i>Verbs</i>	<i>Total</i>
<b>1stOrderEntities</b>	<b>491</b>		<b>491</b>
<b>2ndOrderEntities</b>	<b>272</b>	<b>228</b>	<b>500</b>
<b>3rdOrderEntities</b>	<b>33</b>		<b>33</b>
<b>Total</b>	<b>796</b>	<b>228</b>	<b>1024</b>



## Starting points for the EuroWordNet Top-Ontology

### **Purpose:**

- a) *The ontology should support the building and encoding of semantic networks as linguistic ontologies: networks of lexicalized words and expressions in a language.*
- b) *The classification of the Base Concepts in terms of the Top Ontology should apply to all the involved languages.*
- c) *Enforce uniformity and compatibility of the different wordnets, by providing a common framework. Divide the Base Concepts (BCs) into coherent clusters to enable contrastive-analysis and discussion of closely related word meanings*
- d) *Customize the database by assigning features to the top-concepts, irrespective of language-specific structures.*
- e) *Provide an anchor point for connecting other ontologies to the Inter-Lingual-Index, such as CYC, MikroKosmos, the Upper-Model, by linking them to the corresponding ILLI-records.*

---

## Principles for deciding on the relevant distinctions.

### Starting point is that the wordnets are linguistic ontologies:

1. *Semantic classifications common in linguistic paradigms: Aktionsart models [Vendler 1967, Verkuyl 1972, Verkuyl 1989, Pustejovsky 1991], entity-orders [Lyons 1977], Aristotle's Qualia-structure [Pustejovsky 1995].*
2. *Ontologies developed in previous EC-projects, which had a similar basis and are well-known in the project consortium: Acquilex (BRA 3030, 7315), Sift (LE-62030, [Vossen and Bon 1996]).*
3. *The ontology should be capable of reflecting the diversity of the set of common BCs, across the 4 languages. In this sense the classification of the common BCs in terms of the top-concepts should result in:*
  - *Homogeneous Base Concept Clusters: classifications in WordNet1.5 and the other wordnets.*
  - *Average-sized Base Concept Clusters: not extremely large or small.*

## Other important characteristics:

- *The distinctions apply to both nouns, verbs and adjectives, because these can be related in the language-specific wordnets via a `xpos_synonymy` relation, and the ILI-records can be related to any part-of-speech.*
- *The top-concepts are hierarchically ordered by means of a subsumption relation but there can only be one super-type linked to each top-concept: multiple inheritance between top-concepts is not allowed.*
- *In addition to the subsumption relation top-concepts can have an opposition-relation to indicate that certain distinctions are disjunct, whereas others may overlap.*
- *There may be multiple relations from ILI-records to top-concepts: the Base Concepts can be cross-classified in terms of multiple top-concepts (as long as these have no opposition-relation between them): i.e. multiple inheritance from Top-Concept to Base Concept is allowed.*

**Result:** *the TCs function as cross-classifying features rather than conceptual classes.*

Meanings for *bodyparts* are not linked to a single class **BodyPart** but to two features: **Living** and **Part**.

---

# The EuroWordNet Top-Ontology

**63 concepts (excluding the top)**

## First Level [Lyons 1977]:

### **1stOrderEntity** (*491 BC synsets, all nouns*)

Any concrete entity (publicly) perceivable by the senses and located at any point in time, in a three-dimensional space.

### **2ndOrderEntity** (*500 BC synsets, 272 nouns and 228 verbs*)

Any Static Situation (property, relation) or Dynamic Situation, which cannot be grasped, heard, seen, felt as an independent physical thing. They can be located in time and occur or take place rather than exist; e.g. continue, occur, apply

### **3rdOrderEntity** (*33 BC synsets, all nouns*)

An unobservable proposition that exists independently of time and space. They can be true or false rather than real. They can be asserted or denied, remembered or forgotten. E.g. idea, thought, information, theory, plan.

---

Tests to distinguish between 1st, 2nd and 3rd order entities:

- a The same person was here again to-day
- b The same thing happened/occurred again to-day
- \*? The idea, fact, expectation, etc.... was here/occurred/ took place

Third-order entities cannot occur, have no temporal duration and therefore fail on both tests. A positive test for a 3rdOrderEntity is based on the properties that can be predicated:

- ok The idea, fact, expectation, etc.. is true, is denied, forgotten

The first division of the ontology is disjoint: BCs cannot be classified as combinations of these TCs. This distinction cuts across the different parts of speech in that:

- *1stOrderEntities are always (concrete) nouns.*
- *2ndOrderEntities can be nouns, verbs and adjectives, where adjectives are always non-dynamic (refer to states and situations not involving a change of state).*
- *3rdOrderEntities are always (abstract) nouns.*

### **Base Concepts classified as 3rdOrderEntities:**

theory; idea; structure; evidence; procedure; doctrine; policy; data point; content; plan of action; concept; plan; communication; knowledge base; cognitive content; know-how; category; information; abstract; info;

1stOrderEntity	
<b>Origin</b> Natural Living Plant Human Creature Animal Artifact	<b>Function</b> Vehicle Representation MoneyRepresentation LanguageRepresentation Image Representation Garment Furniture Covering Occupation Software Place Instrument Building Comestible Container
<b>Form</b> Substance Solid Liquid Gas Object	<b>Composition</b> Part Group

- a) **Origin:** *the way in which an entity has come about.*
- b) **Form:** *as an a-morf substance or as an object with a fixed shape, hence the subdivisions Substance and Object.*
- c) **Composition:** *as a group of self-contained wholes or as a part of such a whole, hence the subdivisions Part and Group.*
- d) **Function:** *the typical activity or action that is associated with an entity.*

## Conjunctive classes of 1stOrderEntities

5	Comestible;Artifact		
5	Comestible;Solid;Artifact		
5	Container;Part;Solid;Living		
5	Furniture;Object;Artifact	<i>fruit:</i>	<i>Comestible (Function)</i>
5	Instrument;Artifact		<i>Object (Form)</i>
5	Living		<i>Part (Composition)</i>
5	Plant		<i>Plant (Natural, Origin)</i>
6	Liquid	<i>skin:</i>	<i>Covering (Covering)</i>
6	Object;Artifact		<i>Solid (Form)</i>
6	Part;Living		<i>Part (Composition)</i>
6	Place;Part;Solid		<i>Living (Natural, Origin)</i>
7	Building;Object;Artifact	<i>life:</i>	<i>Group (Composition)</i>
7	Group		<i>Living (Natural, Origin)</i>
7	LanguageRepresentation	<i>cell:</i>	<i>Part (Composition)</i>
7	Vehicle;Object;Artifact		<i>Living (Natural, Origin)</i>
10	Instrument;Object;Artifact	<i>arms:</i>	<i>Instrument (Function)</i>
12	Part		<i>Group (Composition)</i>
14	Place		<i>Object (Form)</i>
14	Place;Part		<i>Artifact (Origin)</i>
15	Substance		
19	LanguageRepresentation;Artifact		
20	Occupation;Object;Human		
22	Object;Animal		
26	Function		
38	Group;Human		
42	Object;Human		

---

## 1stOrderEntities classified as Function only

asset 2	decoration 2	possession 1
barrier 1	device 4	protection 4
belonging 2	fastener 1	remains 2
building material 1	force 6	restraint 2
causal agency 1	force 7	support 6
commodity 1	form 5	support 7
consumer goods 1	impediment 1	supporting structure 1
creation 3	medicament 1	thing 3
curative 1	piece of work 1	



2ndOrderEntity		
SituationType	SituationComponent	
Dynamic <i>(he sat down quickly.            a quick, wild meeting)</i> BoundedEvent UnboundedEvent	Cause Agentive Phenomenal Stimulating Communication Condition Existence Experience Location Manner	Mental Modal Physical Possession Purpose Quantity Social Time Usage
Static <i>(?he sits quickly.)</i> Property Relation		

- **SituationType:** *the event-structure in terms of which a situation can be characterized as a conceptual unit over time; Disjoint features*
- **SituationComponent:** *the most salient semantic component(s) that characterize(s) a situation; Conjoined Features.*

## Conjunctive classes of 2ndOrderEntities

### Static

5	Property;Physical;Condition
5	Property;Stimulating;Physical
5	Relation
5	Relation;Social
6	Static

6	Static;Quantity
7	Property;Condition
8	Relation;Location
9	Property
10	Relation;Physical;Location

**Relation;Physical;Location:** adjoin 1; aim 4; blank space 1; course 7; direction 8; distance 1; elbow room 1; path 3; spatial property 1; spatial relation 1

### Dynamic

5	BoundedEvent;Agentive;Purpose;Communication; Social
5	BoundedEvent;Cause;Physical
5	BoundedEvent;Cause;Physical;Location
5	BoundedEvent;Time
5	Dynamic
5	Dynamic;Location
5	Dynamic;Phenomenal
5	Dynamic;Phenomenal;Physical
6	BoundedEvent;Agentive

6	BoundedEvent;Location
6	BoundedEvent;Physical;Location
6	Dynamic;Agentive;Communication
6	Dynamic;Cause
6	UnboundedEvent;Agentive;Purpose;Social
8	BoundedEvent;Agentive;Mental;Purpose
8	BoundedEvent;Quantity;Time
9	BoundedEvent;Cause
9	Dynamic;Experience;Mental

**Dynamic;Experience;Mental:** experience 7; find 3; affect 5; arouse 5; excite 2; cognition 1; desire 2; disposition 2; disposition 4; disturbance 7; emotion 1; feeling 1; humor 3; pleasance 1; process 4; look 8; phenomenon 1; cause to appear 1; perception 2; sensation 1; feel 12; experience 8; trouble 3; reality 1

## **Adapting the Inter-Lingual-Index:**

1. Extending the Inter-Lingual-Index to be the superset of all concepts occurring in the separate wordnets.
  
2. Globalizing the sense-differentiation across the resources.

## Extending the Inter-Lingual-Index to be the superset of all concepts.

### Motivation:

- *It should be possible to link equivalent non-English meanings (e.g. Italian-Spanish) to the same ILI-record even when there is no English or WordNet equivalent.*
- *It should be possible to store domain-labels, glosses or top-concept classifications for non-English meanings, e.g.: all Spanish bull-fighting terms should be linked to ILI-records with the domain-label bull-fighting.*

### Procedure:

1. *Initially, the ILI will only contain WordNet1.5 synsets.*
2. *a site that cannot find a proper equivalent among the available ILI-concepts will link the meaning to another ILI-record using a so-called complex-equivalence relation and will generate a potential new ILI-record:*

<b>Dutch Meaning</b>	<b>Definition</b>	<b>Complex-equivalence</b>	<b>Target concept</b>
klunen	to walk on skates	HAS_EQ_HYPERONYM	walk

3. *after a building-phase all potentially-new ILI-records are collected and verified for overlap by one site;*
4. *a proposal for updating the ILI is distributed to all sites and has to be verified;*
5. *the ILI is updated and all sites have to reconsider the equivalence relations for all meanings that can potentially be linked to the new ILI-records;*

## Globalize the sense-differentiation across the resources.

### Globalized meaning for “clean”, grouping 4 out of 19 verb meanings:

ILI-ID	@62475@
Synset	<b>clean, remove unwanted substances from, pick, make clean, scavenge</b>
Source-references	
Source-id	106221: <b>(make clean by removing dirt, filth, or unwanted substances from; "Clean the stove!"; "The dentist cleaned my teeth");</b>
Word form	Clean, VERB, @61238@
Source-id	109110 <b>(remove unwanted substances from, such as feathers or pits, as of chickens or fruit; "Clean the turkey");</b>
Word form	Clean, VERB, @61262@
Source-id	881863 <b>(remove in making clean; "Clean the spots off the rug");</b>
Word form	Clean, VERB, @66910@
Source-id	881979 <b>(as in chemistry)</b>
Word form	Clean, VERB, @66911@
Polysemy-type	Generalization

Dutch WordNet	Spanish WordNet
schoonmaken 1	limpiar 2
ILI-reference	ILI-reference
Eq_near_synonym @61262@	Eq_synonym @61238@
Eq_near_synonym @66911@	
Eq_generalization @62475@	Eq_generalization @62475@

**Metonymy-relation between specific meaning of “university”:**

ILI-ID	@62489@
Synset	University
Source-references	
Source-id	2039764 ( <b>where a seat of higher learning is housed, including administrative and living quarters as well as facilities for research and teaching</b> );
Word form	University, NOUN, @12547@
Source-id	5276749 ( <b>the faculty and students of a university</b> );
Word form	University, NOUN, @35629@
Polysemy-type	Metonymy

<b>Dutch WordNet</b>	<b>Spanish WordNet</b>	<b>Italian WordNet</b>
<i>universiteit</i> 1 {institution}	<i>universidad</i> 1 {institution; building}	<i>universitare</i> 1 {building}
ILI-reference	ILI-reference	ILI-reference
Eq_synonym @35629@	Eq_near_synonym @12547@	Eq_synonym @12547@
Eq_metonym @62489@	Eq_near_synonym @35629@	Eq_metonym @62489@
<i>universiteit</i> 2 {building}	Eq_metonym @62489@	
ILI-reference		
Eq_synonym @12547@		
Eq_metonym @62489@		