# Ontology as Interlingua

# Crosslinguistic WordNets

Starting in late 1990s, WordNets were built for languages other than English

Genetically and typologically unrelated languages: Turkish, Hindi, Chinese, Korean, Basque, Xhosa, Arabic, Latin… (currently >70)

http://www.globalwordnet.org

# Wordnets

Entire families of wordnets: EuroWordNet, BalkaNet, IndoWordNet, AsianWordNet, AfricanWordNet,…

Local product: MultiWordNet

# Wordnets in the world

Motivations:

- Natural Language Processing applications that require word sense discrimination and disambiguation within and across languages
- Crosslingual comparison of lexical categories
- Interesting by-product: discover language-specific lexical gaps

# What is universal?

- Surely not all “concepts”:
- English has many verbs of walking (*slouch, strut, stroll,amble, prance, sneak, march*,..) and walking/running (*hop, skip, bounce*,…)
- No 1:1 crosslingual encoding of concepts
- But is the network structure universal? Can all words in all languages be connected?
- Are the relations universal (if so, this would strengthen their cognitive reality)

# Classes

- Crosslingual construction reveals potentially meaningful classes
- Classifiers (Chinese, Bantu)
- Represent roots as underspecified semantic categories separately from words in semitic languages

# Crosslinguistic WordNets

Some are manually constructed

--independently from PWN, mapped later ("Merge" method)

or

--translated directly from PWN ("Expand" method)

First method is considered easier, more accurate (why?)

Other wordnets are constructed semi-automatically

# Crosslinguistic Wordnets

All new wordnets are mapped to the Princeton WordNet , which serves as a general model and as the link for connecting the wordnets

# Mapping words and synsets across multilingual WordNets

First set of eight foreign-language WNs (EuroWordNet; Vossen 1998) were built with reference to Princeton WordNet

Princeton WN as the hub ("interlingual index")

Each synset in each WN was linked to a "record" (PWN synset identifier) in the index

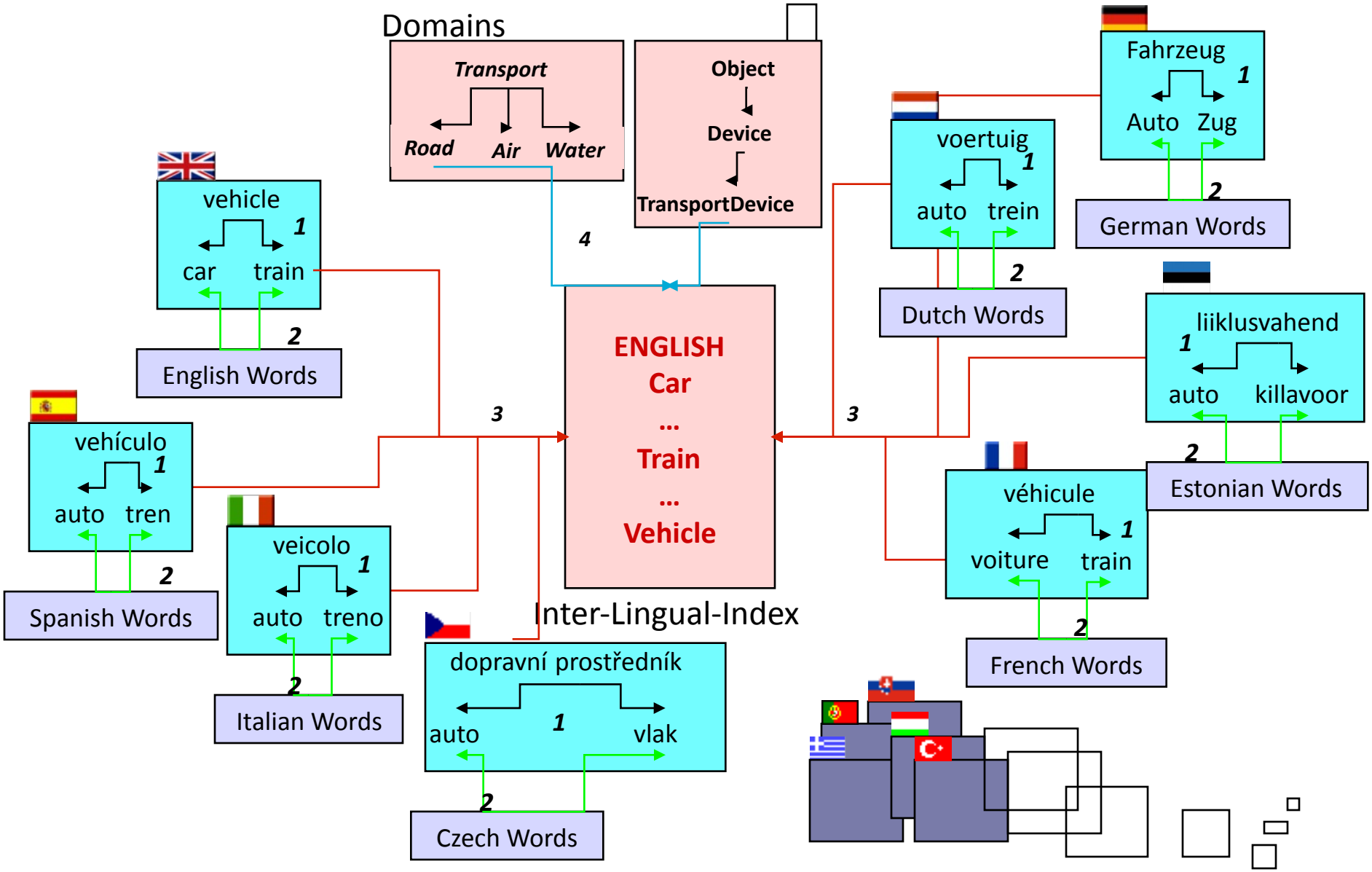
Crosslingual mapping of words and synsets proceeds via the index

# Mapping words and synsets across multilingual WordNets

The Interlingual Index is a flat, unstructured list

Princeton WN's **structure** is not imported

Only the language-specific wordnets have relations and form networks


Domains
Transport
Road
Air
Water
Object
Device
TransportDevice
4
ENGLISH
Car
...
Train
...
Vehicle
Inter-Lingual-Index
3
3
vehicle
1
car
train
2
English Words
vehículo
1
auto
tren
2
Spanish Words
veicolo
1
auto
treno
2
Italian Words
dopravní prostředník
auto
1
vlak
2
Czech Words
voertuig
1
auto
trein
2
Dutch Words
Fahrzeug
1
Auto
Zug
2
German Words
liiklusvahend
1
auto
killavoor
2
Estonian Words
véhicule
1
voiture
train
2
French Words

# Mismatches in multilingual WordNets

Concepts not lexicalized in English required the creation of new records in the ILI (w/out English synsets or synsets in some other wordnets)

E.g., Arabic lexically distinguishes more kinds of *cousin* than English; thus the ILI needs appropriate placeholders (records)

Xhosa time expressions:

*the_time_of_day_when_you_are_beautiful*

*the_time_of_day_when_you_see_the_horns_of_the_cattle_against-the_sky*

# Mismatches in multilingual WordNets

Conversely, some languages lack the equivalents of English words:

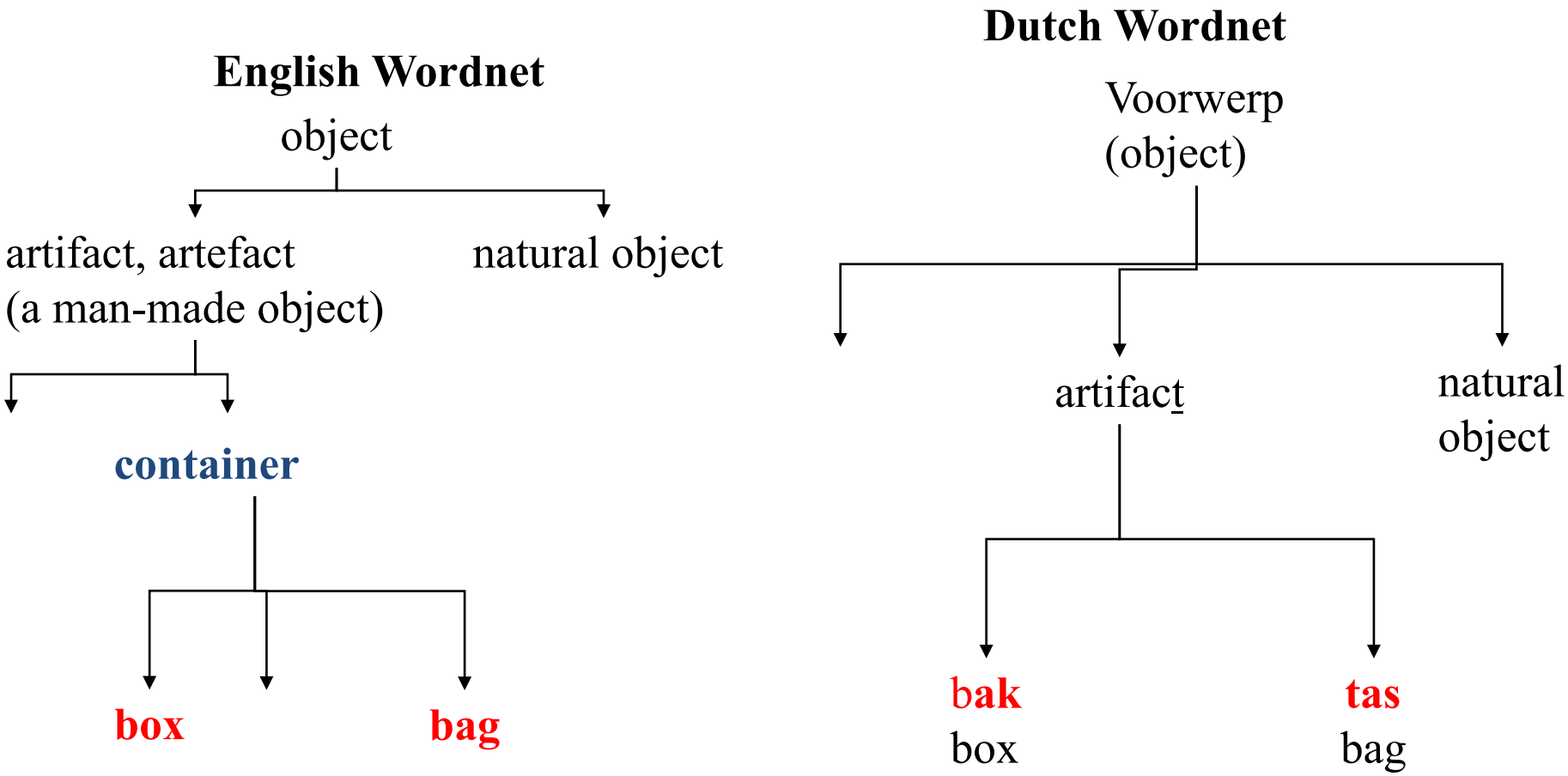
--Dutch lacks *container* but lexicalizes kinds (hyponyms) of *container* (*box, bag, bottle, bowl...*)

Respective hierarchies reflects this difference; Dutch wordnet "skips" a level

Du. *bag, box...=>artifact*

Engl. *bag, box…=>container =>artifact*

# English-Dutch snippet



# Multilingual WordNets

Interlingual Index in EuroWordNet is biased towards English

Could skew coverage of new wordnets, esp. those that are translated from English

Some mapped synsets aren't really equivalent

# Interlingua

Solution: replace index by *language-independent, formal* ontology that can accommodate the lexicons of all languages

Meanings are stated as axioms in logical form

Axioms are machine-readable

Interlingual ontology enables automatic reasoning and inferencing, within and across languages

# Some proposed ontologies

- SUMO
- DOLCE
- KYOTO