WordNet
EuroWordNet, and Global WordNet

Mustafa Jarrar

Sina Institute, University of Birzeit

mjarrar@birzeit.edu
www.jarrar.info

Watch this lecture and download the slides from

http://jarrar-courses.blogspot.com/2011/11/artificial-intelligence-fall-2011.html
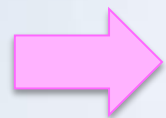
Reading

Everything in these slides   +   everything I say

[MBC93] George A. Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine Miller: **Introduction to WordNet: An On-line Lexical Database**. International Journal of Lexicography, Vol. 3, Nr. 4. Pages 235-244. (1990) http://wordnetcode.princeton.edu/5papers.pdf

[GGO02] Aldo Gangemi , Nicola Guarino , Alessandro Oltramari , Ro Oltramari , Stefano Borgo: **Cleaning-up WordNet's Top-Level**. In Proc. of the 1st International WordNetConference (2002)

http://citeseer.ist.psu.edu/viewdoc/download;jsessionid=C9962DFEDD793F3F839426B774BC9BAF?doi=10.1.1.11.4064&rep=rep1&type=pdf

WordNet and Global WordNet

➡️ **Part 1: The English WordNet**

- **Part 2: Euro WordNet**

- **Part 3: Global WordNet**

Lecture Keywords:

WordNet, Global WordNet, Thesaurus, Linguistic Ontology, Lexical Semantics, Semantics, Meaning, Synset, Concept, Synonymy, Polysemy, Hyponymy, Meronymy, Antonymy,
شبكة المفردات، مكنز، انطولوجيا للغة، علم الدلالة، الدلالة، المعنى، المفهوم،
تعدد اللغات، الترادف اللغوي، تعدد المعاني، التضاد، تصنيف المعاني، علاقات
جزء-كل

What is WordNet?

- In 1985 a group of **psychologists and linguists** at Princeton University started to develop a "**mental lexicon**".

- You may also call it:"electronic dictionary", "Mental dictionary", English, "**semantic Network**", hyperdimensional thesaurus, etc.

- Includes **most frequent words** (nouns, adjectives, adverbs, verbs).

- **Organized by meaning**: words in close proximity are semantically similar.

- Can be used by humans and machines.

- Human users and computers can browse WordNet and find words that are meaningfully related to their queries.

- **Available online**, for downloading! http://wordnet.princeton.edu

WordNet: **Synonymy**

WordNet gives information about two fundamental, universal properties of human language: **polysemy** and **synonymy**.

- English words are grouped (roughly) into sets of synonyms.

- Each set of synonyms is called a **Synset**; and given a unique **SynsetID** to identify it.

- Each synset expresses a distinct meaning/**concept**.

03410635

{**Furniture, Piece of furniture , Article of furniture**}

Furnishings that make a room….

03018908

{**Bureau, Dresser, Chest of Drawers,**}

Furniture with drawers for keeping clothes

07955878

{**Categorization, Classification**}

A group of people or things arranged…

08283156

{**Table**, **Tabular Array**}

A set of data arranged in rows and columns

04615793

{**work table**}

A table designed…

06501650

{**Contents, TableOfContents**}

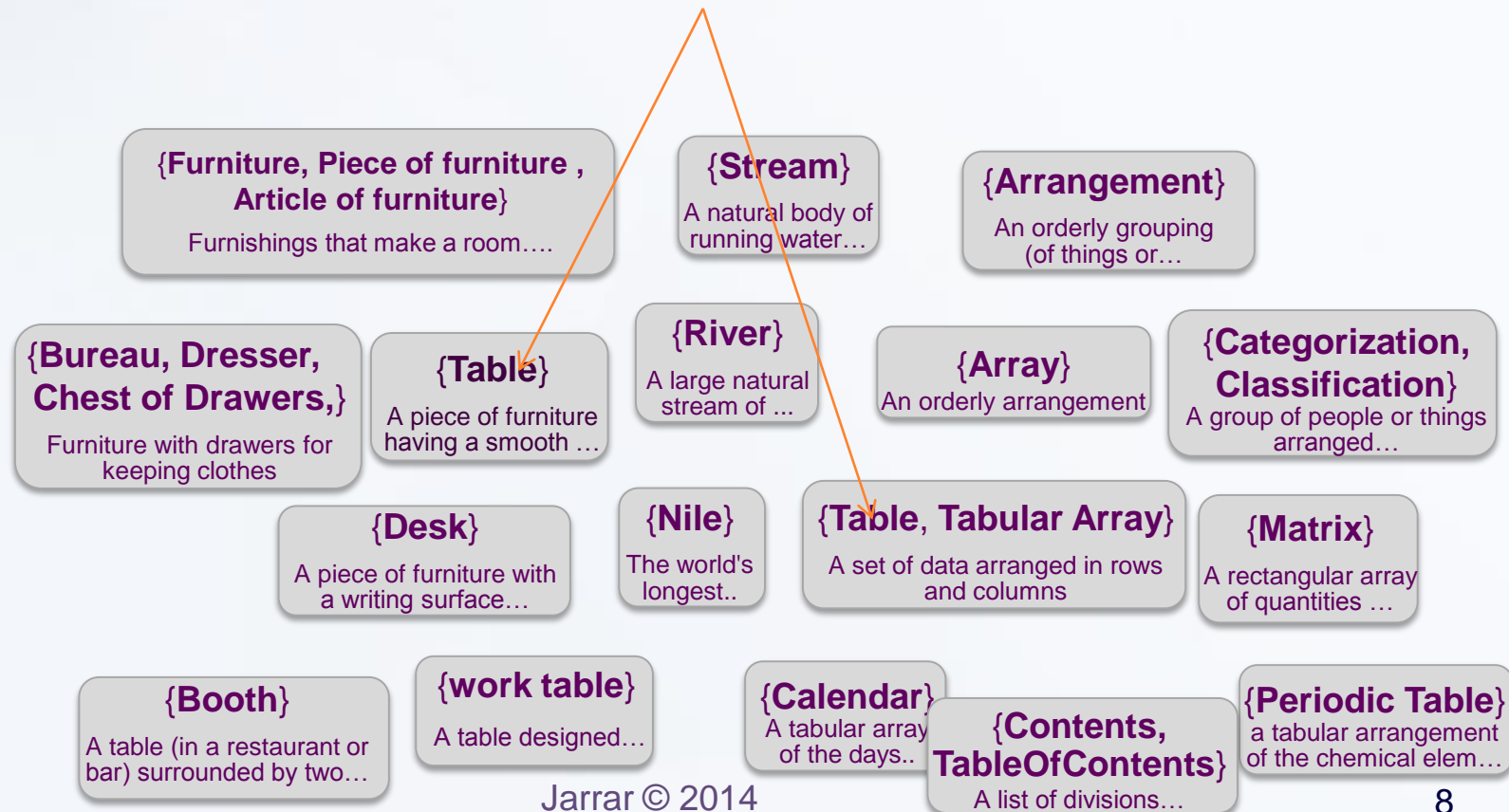A list of divisions…

Jarrar © 2014

6

Exercise

List the different meanings of the words:

Table,   Array,   Matrix,   Bureau

WordNet: **Polysemy**

- Each word form-meaning pair is unique.

- A word that appears in *n* synsets is *n*-fold polysemous.

- For example: "Table" here is two-fold polysemous

{**Furniture, Piece of furniture , Article of furniture**}

Furnishings that make a room….

{**Stream**}

A natural body of running water…

{**Arrangement**}

An orderly grouping (of things or…

{**Bureau, Dresser, Chest of Drawers,**}

Furniture with drawers for keeping clothes

{**Table**}

A piece of furniture having a smooth …

{**River**}

A large natural stream of ...

{**Array**}

An orderly arrangement

{**Categorization, Classification**}

A group of people or things arranged…

{**Desk**}

A piece of furniture with a writing surface…

{**Nile**}

The world's longest..

{**Table, Tabular Array**}

A set of data arranged in rows and columns

{**Matrix**}

A rectangular array of quantities …

{**Booth**}

A table (in a restaurant or bar) surrounded by two…

{**work table**}

A table designed…

{**Calendar**}

A tabular array of the days..

{**Contents, TableOfContents**}

A list of divisions…

{**Periodic Table**}

a tabular arrangement of the chemical elem…

WordNet: **Glosses**

A short gloss is provided for each sysnet.

Glosses are examples of contexts for many word-sense pairs, telling us how words with specific senses are being used in context.

{**Furniture, Piece of furniture , Article of furniture**}

Furnishings that make a room….

{**Stream**}

A natural body of running water…

{**Arrangement**}

An orderly grouping (of things or…

{**Bureau, Dresser, Chest of Drawers,**}

Furniture with drawers for keeping clothes

{**Table**}

A piece of furniture having a smooth …

{**River**}

A large natural stream of ...

{**Array**}

An orderly arrangement

{**Categorization, Classification**}

A group of people or things arranged…

{**Desk**}

A piece of furniture with a writing surface…

{**Nile**}

The world's longest..

{**Table, Tabular Array**}

A set of data arranged in rows and columns

{**Matrix**}

A rectangular array of quantities …

{**Booth**}

A table (in a restaurant or bar) surrounded by two…

{**work table**}

A table designed…

{**Calendar**}

A tabular array of the days..

{**Contents, TableOfContents**}

A list of divisions…

{**Periodic Table**}

a tabular arrangement of the chemical elem…

WordNet: **Statistics**

155 287 word forms, groups into

117 659 synsets

	WordForms	Synsets
noun	117,798	82,115
verb	11,529	13,767
adjective	21,479	18,156
adverb	4,481	3,621
Total	**155,287**	**117,659**

{**Furniture, Piece of furniture , Article of furniture**}
Furnishings that make a room….

{**Stream**}
A natural body of running water…

{**Arrangement**}
An orderly grouping (of things or…

{**Bureau, Dresser, Chest of Drawers,**}
Furniture with drawers for keeping clothes

{**Table**}
A piece of furniture having a smooth …

{**River**}
A large natural stream of ...

{**Array**}
An orderly arrangement

{**Categorization, Classification**}
A group of people or things arranged…

{**Desk**}
A piece of furniture with a writing surface…

{**Nile**}
The world's longest..

{**Table, Tabular Array**}
A set of data arranged in rows and columns

{**Matrix**}
A rectangular array of quantities …

{**Booth**}
A table (in a restaurant or bar) surrounded by two…

{**work table**}
A table designed…

{**Calendar**}
A tabular array of the days..

{**Contents, TableOfContents**}
A list of divisions…

{**Periodic Table**}
a tabular arrangement of the chemical elem…

WordNet Semantic Relations

Synsets are interconnected with semantic relations, forming a large semantic network (graph).
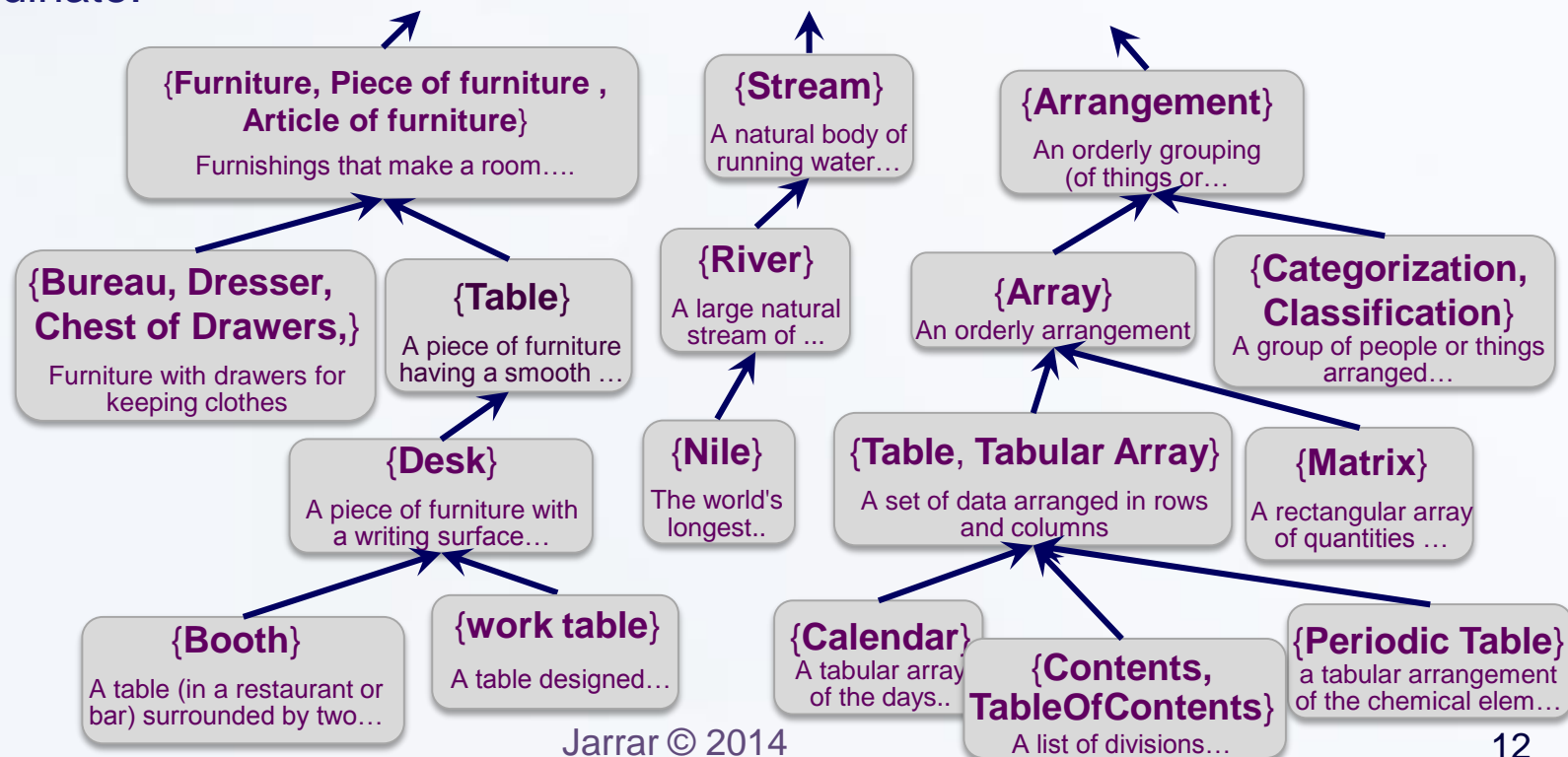
Such Relations are:
- **Hyponymy**, also called "Is a" relation, or sub/superordinate.
- **Meronymy**, also called "part of" relation

{Container}
Any object that can be used ..

{Drawer}
A boxlike container in a..

{shelf}
A support that consists…

{Support}
Any device that bears..

{Furniture, Piece of furniture , Article of furniture}
Furnishings that make a room….

{Bureau, Dresser, Chest of Drawers,}
Furniture with drawers for keeping clothes

{Table}
A piece of furniture having a smooth …

{Desk}
A piece of furniture with a writing surface…

{Booth}
A table (in a restaurant or bar) surrounded by two…

{work table}
A table designed…

{Stream}
A natural body of running water…

{River}
A large natural stream of ...

{Nile}
The world's longest..

{Arrangement}
An orderly grouping (of things or…

{Array}
An orderly arrangement

{Categorization, Classification}
A group of people or things arranged…

{Table, Tabular Array}
A set of data arranged in rows and columns

{Matrix}
A rectangular array of quantities …

{Calendar}
A tabular array of the days..

{Contents, TableOfContents}
A list of divisions…

{Periodic Table}
a tabular arrangement of the chemical elem…

WordNet Relations: **Hyponymy**

- A synset {x, x′, . . .} is hyponym of the synset {y, y′, . . .} **if native English speakers accept sentences like x is a (kind of) y**. E. g., *Table/Tabular Array* is a kind of *Array*, *Array* is a kind of *Arrangement*,…

- Hyponymy is transitive and asymmetrical. So as Hyponymy generates a hierarchical semantic structure, a hyponym inherits all the features of the more generic concept and adds at least one feature that distinguishes it from its superordinate.

{**Furniture, Piece of furniture , Article of furniture**}
Furnishings that make a room….

{**Stream**}
A natural body of running water…

{**Arrangement**}
An orderly grouping (of things or…

{**Bureau, Dresser, Chest of Drawers,**}
Furniture with drawers for keeping clothes

{**Table**}
A piece of furniture having a smooth …

{**River**}
A large natural stream of ...

{**Array**}
An orderly arrangement

{**Categorization, Classification**}
A group of people or things arranged…

{**Desk**}
A piece of furniture with a writing surface…

{**Nile**}
The world's longest..

{**Table, Tabular Array**}
A set of data arranged in rows and columns

{**Matrix**}
A rectangular array of quantities …

{**Booth**}
A table (in a restaurant or bar) surrounded by two…

{**work table**}
A table designed…

{**Calendar**}
A tabular array of the days..

{**Contents, TableOfContents**}
A list of divisions…

{**Periodic Table**}
a tabular arrangement of the chemical elem…

Jarrar © 2014

12

WordNet Relations: **Hyponymy**

- A synset {x, x′, . . .} is hyponym of the synset {y, y′, . . .} **if native English speakers accept sentences like x is a (kind of) y.** E. g., *Table/Tabular Array* is a kind of *Array*, *Array* is a kind of *Arrangement*,…

- Hyponymy is transitive and asymmetrical. So as Hyponymy generates a hierarchical semantic structure, a hyponym inherits all the features of the more generic concept and adds at least one feature that distinguishes it from its superordinate. [2]

The WordNet hierarchy is about 16 levels

Top Level Nouns (25 unique beginners)

{act, action, activity}	{natural object }
{animal, fauna}	{natural phenomenon }
{artifact }	{person, human being}
{attribute, property }	{plant, flora}
{body, corpus}	{possession}
{cognition, knowledge}	{process}
{communication}	{quantity, amount}
{event, happening}	{relation }
{feeling, emotion}	{shape}
{food}	{state, condition}
{group, collection}	{substance}
{location, place }	{time}
{motive}	

WordNet Relations: **Meronymy**

- A synset {x, x′, . . .} is meronym of the synset {y, y′, . . .} if native English speakers accept sentences like y has an x (as a part) or An x is a part of y. E. g., *Finger* is part of *Hand* , *Hand* is part of *Arm, Arm* is part of *Body*.

- Meronymy is transitive (with qualification) and asymmetrical relations, and forms a part hierarchy..

- Synsets may have multiple hypernyms

{Container}
Any object that can be used ..

{Drawer}
A boxlike container in a..

{shelf}
A support that consists…

{Support}
Any device that bears..

{Furniture, Piece of furniture , Article of furniture}
Furnishings that make a room….

{Bureau, Dresser, Chest of Drawers,}
Furniture with drawers for keeping clothes

{Table}
A piece of furniture having a smooth …

{Desk}
A piece of furniture with a writing surface…

{Booth}
A table (in a restaurant or bar) surrounded by two…

{work table}
A table designed…

{Stream}
A natural body of running water…

{River}
A large natural stream of ...

{Nile}
The world's longest..

{Arrangement}
An orderly grouping (of things or…

{Array}
An orderly arrangement

{Categorization, Classification}
A group of people or things arranged…

{Table, Tabular Array}
A set of data arranged in rows and columns

{Matrix}
A rectangular array of quantities …

{Calendar}
A tabular array of the days..

{Contents, TableOfContents}
A list of divisions…

{Periodic Table}
a tabular arrangement of the chemical elem…

Jarrar © 2014

14

Exercise

Find the hyponyms and meronyms of this synset

{car, auto, automobile, machine, motorcar}

WordNet Relations: **Another Example**

Hyponymy and meronymy relations are:
* transitive
* directed

{conveyance,transport}

{vehicle}

meronyms

{motor vehicle, automotive vehicle}

hyper(o)nym

{car, auto, automobile, machine, motorcar}

hyponym

{cruiser, squad car, patrol car, police car, prowl car}

{cab, taxi, hack, taxicab}

{car mirror}

{car door}

{bumper}

{car window}

{armrest}

{doorlock}

{hinge, flexible joint}
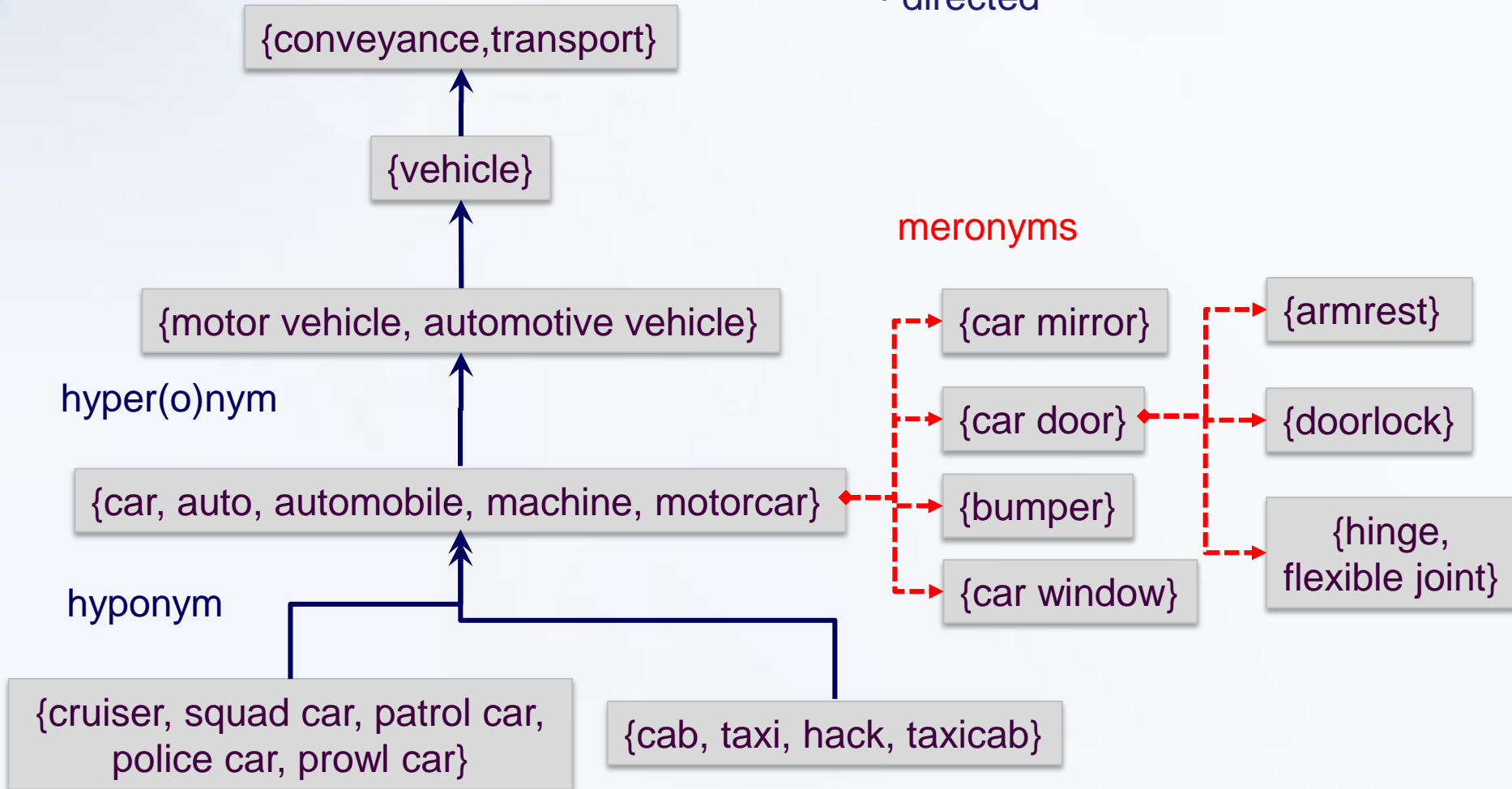
WordNet Relations: **Antonymy**

- The antonym of a word x is sometimes not-x, but not always. For example, *rich* and *poor* are antonyms, but to say that someone is not rich does not imply that they must be poor; many people consider themselves neither rich nor poor.

- Antonymy, which seems to be a simple symmetric relation, is actually quite complex, yet speakers of English have little difficulty recognizing antonyms when they see them. For example, the meanings {rise, ascend } and {fall, descend} may be conceptual opposites, but they are not antonyms; [rise/fall] are antonyms and so are [ascend/descend], but most people hesitate and look thoughtful when asked if rise and descend, or ascend and fall, are antonyms

- Antonymy is a lexical relation between word forms, not a semantic relation between word meanings. Or, some call it semantic relations between words [MPC93].

{Hot}
Used of physical heat; having..

{Cold}
Having a low or inadequate..

{New}
Unaffected by use or exposure

{Worn}
Affected by wear; damaged by …

{New}
Not of long duration; having..

{Old}
Of long duration

{Young, Immature}
in an early period of life…

{Old}
having lived for a relatively

{Rise, Uprise, Come up, Go up, Move up, Lift}
Move upward

{Ascend, Move up, Rise}
Move to a better position in life …

{Ascend, Go up}
Travel up

{Ascend, Come up, Rise, Uprise}
(astronomy) come up, of celestial bodies

{Fall, Come Down, Go Down, Descend}
Move downward and lower, but not necessarily all the way

{Set, Go down, Go Under}
(astronomy) disappear beyond the horizon

Jarrar © 2014

WordWeb

A nice and intuitive interface for WordNet
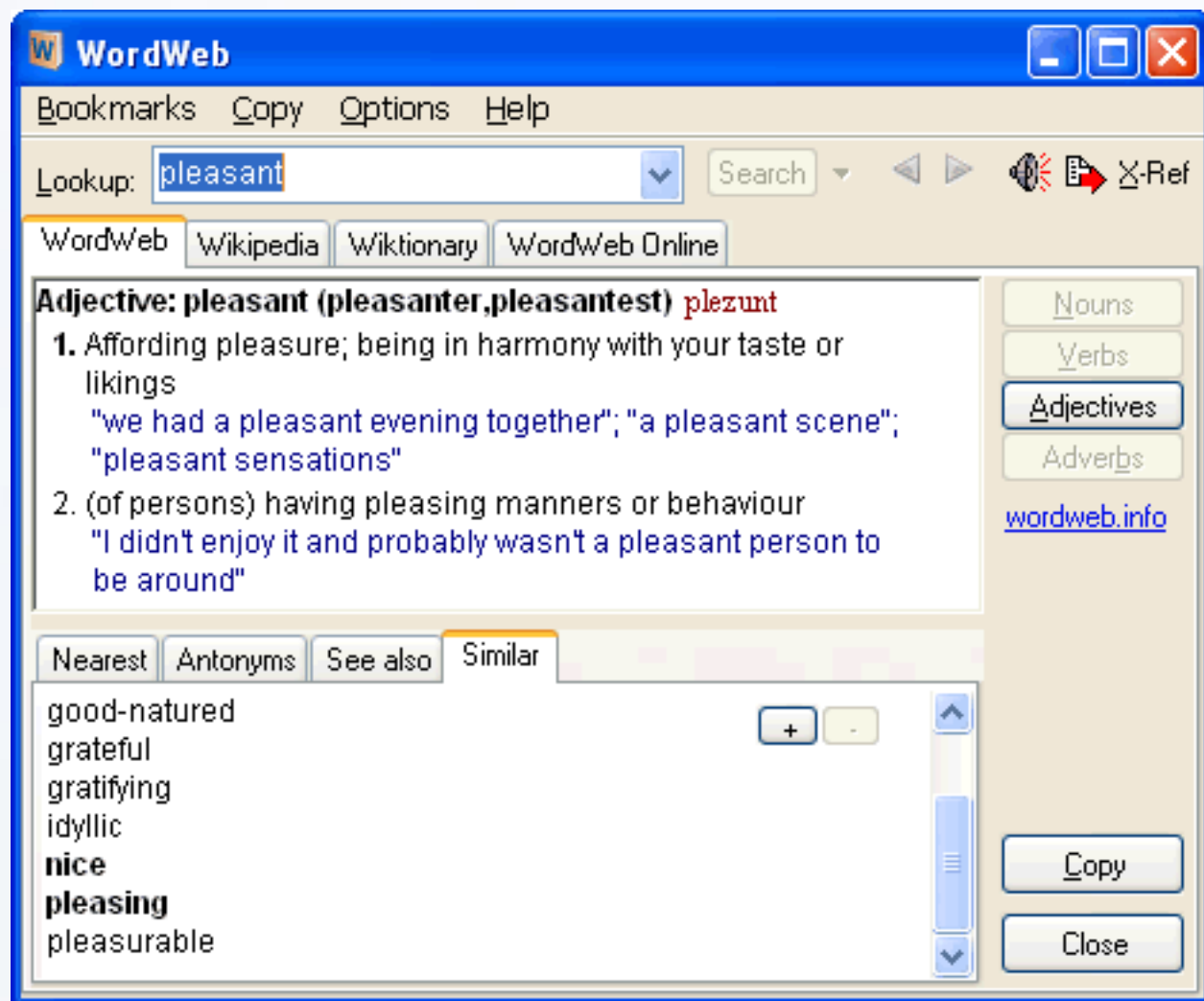
Other WordNet Relations

- Although the main interest of WordNet was on specifying semantic relations but other lexical/morphological relations between word forms were added.

- For example: stems, singular-plural, verb tenses, etc.

Why do we need WordNet?

- Word sense disambiguation,

- Information retrieval,

- Automatic text classification,

- Automatic text summarization,

- Machine translation

- ….etc.

Is WordNet a Thesaurus?

Yes:
- it groups together meaningfully related words

No:
- WN labels the relations

- The relations are limited

- Related words are linked to specific concepts (disambiguated); thesaurus is a "bag of words"

- Many words linked in WordNet do not co-occur in the same thesaurus entry

- WordNet allows one to measure and quantify the semantic similarity or distance among words and concepts

Is WordNet an Ontology?

Meaning (called Ontological Precision):
WordNet: based on what native speakers agree roughly
Ontology: based on Scientific and philosophical findings.

Classification:
WordNet: based on what native speakers agree roughly (Student IsA person)
Ontology: based on strict formal methodologies (student IsA role)

Formal Specification:
WordNet: logically vague
Ontology: strictly formal

→ I like to use WordNet as a linguistic ontology, though it needs lots of cleaning!

→ Linguistic ontologies are difficult to build but they are immune to changes

WordNet and Global WordNet

- **Part 1: The English WordNet**

➡ **Part 2: Euro WordNet**

- **Part 3: Global WordNet**

EURO WordNet

- The development of a multilingual database with WordNets for several European languages.

- Funded by the European Commission, DG XIII, LE2-4003 and LE4-8328
- March 1996 - September 1999   (2.5 Million EURO)
  http://www.hum.uva.nl/~ewn
  http://www.illc.uva.nl/EuroWordNet/finalresults-ewn.html

- **Languages covered:**
  EuroWordNet-1 (LE2-4003): English, Dutch, Spanish, Italian
  EuroWordNet-2 (LE4-8328): German, French, Czech, Estonian.

- **Size of vocabulary:**
  EuroWordNet-1: 30,000 concepts - 50,000 word meanings.
  EuroWordNet-2: 15,000 concepts- 25,000 word meaning.

- **Type of vocabulary:**
  the most frequent words of the languages
  all concepts needed to relate more specific concepts.

EURO WordNet Model

I = Language Independent link
II = Link from Language Specific to Inter lingual Index
III = Language Dependent Link
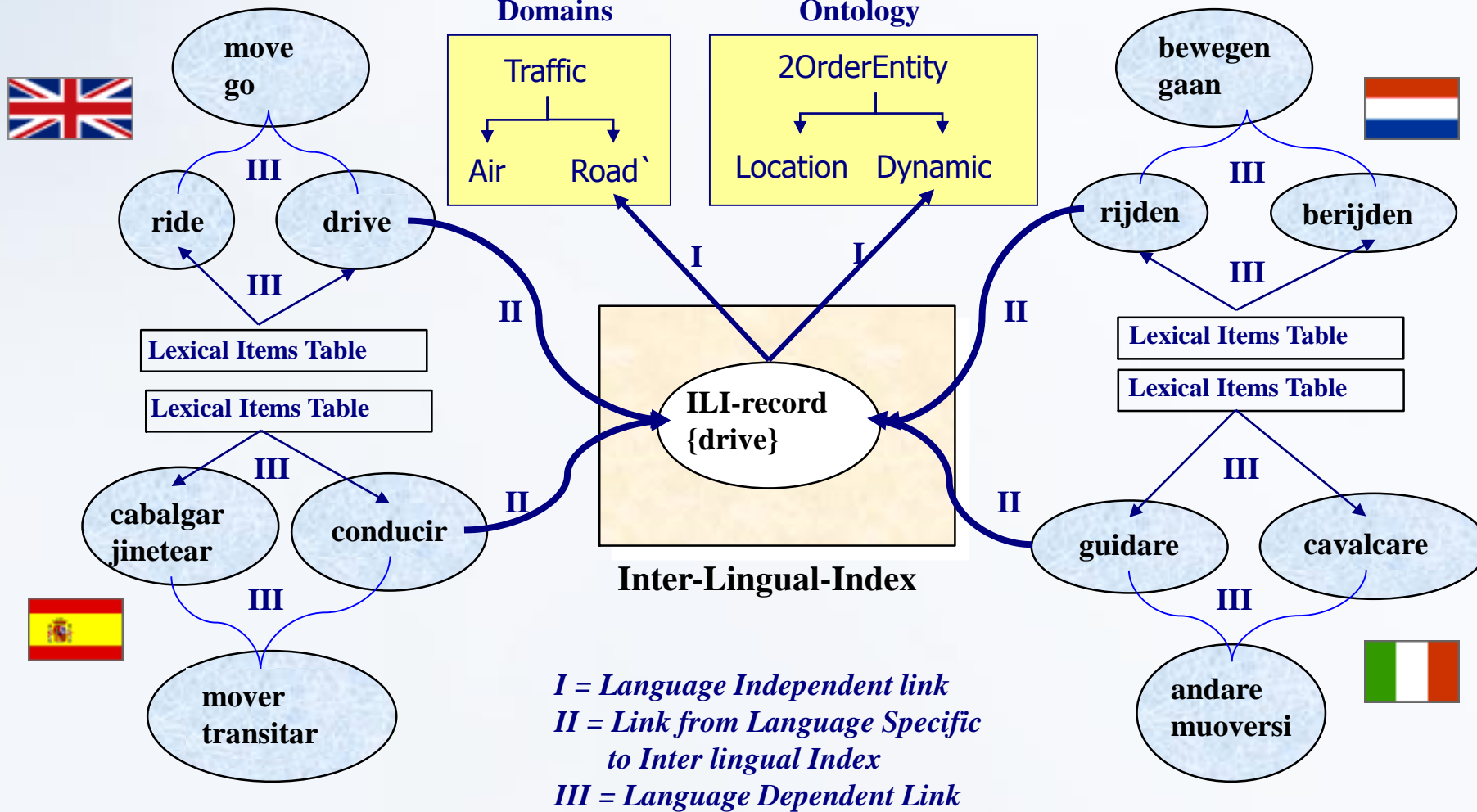
The Multilingual Design

- Inter-Lingual-Index: **unstructured fund of concepts** to provide an efficient mapping across the languages;

- Index-records are mainly **based on WordNet synsets** and consist of synonyms, glosses and source references;

- Various types of **complex equivalence relations** are distinguished;

- Equivalence relations from synsets to index records: **not on a word-to-word basis;**

- **Indirect** matching of synsets linked to the same index items;

EURO WordNet Model

- WordNets are unique language-specific structures:

  - same organizational principles: synset structure and same set of semantic relations.

  - different lexicalizations

  - differences in synonymy and homonymy:

    "decoration" in English versus "versiersel/versiering" in Dutch

    "bank" in English (money/river) versus "bank" in Dutch (money/furniture)

- BUT also different relations for similar synsets

Some Downsides of the EuroWordNet Model

- Construction is not done uniformly

- Coverage differs

- Not all wordnets can communicate with one another, i.e. linked to different versions of English wordnet

- Proprietary rights restrict free access and usage

- A lot of semantics is duplicated

- Complex and obscure equivalence relations due to linguistic differences between English and other languages

WordNet and Global WordNet

- **Part 1: The English WordNet**

- **Part 2: Euro WordNet**

→ **Part 3: Global WordNet**

From EuroWordNet to Global WordNet

http://www.globalwordnet.org
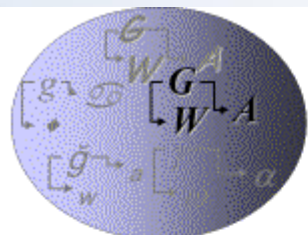
The Global WordNet Association

Global WordNet Association

- *GWC2012 Conference*
- Wordnets in the world
- Wordnet Biblio
- Previous GWA Conferences
- Base Concepts
- The Global Wordnet Grid
- Membership form
- Mailing list
- The Constitution
- The Board
- Background document

The Global WordNet Association is a free, public and non-commercial organization that provides a platform for discussing, sharing and connecting wordnets for all languages in the world. The aims of the association are:

1. To establish distribution facilities for the dissemination of the Association and Association publications and information materials:
   - To promote cooperation and information exchange among related professional and technical societies that build or use wordnets.
   - To provide information on wordnets to the general public.
2. To promote the standardization of the specification of wordnets for all languages in the world, including:
   - the standardization of the Inter-Lingual-Index for inter-linking the wordnets of different languages, as a universal index of meaning
   - the development of a common representation for wordnet data
3. To promote the development of sense-tagged corpora in all the linked languages.
4. To promote sharing and transferring of data, software and specifications across wordnet builders for different languages
5. To promote the development of guidelines and methodologies for building wordnets in new languages
6. To promote the development of explicit criteria and definitions for verifying the relations in any

From EuroWordNet to Global WordNet

- EuroWordNet ended in 1999

- Global Wordnet Association was founded in 2000 to maintain the framework: http://www.globalwordnet.org

- Currently, wordnets exist for more than 50 languages, including:

  Arabic, Bantu, Basque, Chinese, Bulgarian, Estonian, Hebrew, Icelandic, Japanese, Kannada, Korean, Latvian, Nepali, Persian, Romanian, Sanskrit, Tamil, Thai, Turkish, Zulu...

- Many languages are genetically and typologically unrelated

➔ The Arabic WordNet extension was not successful, will be explained later.

Global WordNet Model

Construct separate **wordnets for each language**

Contributors from each language encode the **same core** set of concepts **plus culture/language-specific** ones

Synsets (concepts) are mapped cross linguistically via an ontology instead of just the English Wordnet

Discussion

What would be a good database schema to store WordNet? Global WordNEt?

What is the difference between Synset and Concept?

How precise the Hyponymy? And what is the difference between to Hyponymy and subclass/subset?

References

[1]   Piek Vossen: Lecture Notes on The Global Wordnet Grid: anchoring languages to universal meaning
http://www.authorstream.com/Presentation/Stentore-40555-WN-EWN-GWA-Koszalin-Global-Wordnet-Grid-anchoring-languages-universal-meaning-kosz-Entertainment-ppt-powerpoint/

[2]   Lyons, John. *Semantics*. Vol. 1. Cambridge: Cambridge UP, 1977. Print.