

BabelNet goes to the Multilingual Semantic Web ESWC Tutorial – 26th May 2014

Roberto Navigli

navigli@di.uniroma1.it

David Jurgens

jurgens@di.uniroma1.it

DIPARTIMENTO
DI INFORMATICA



SAPIENZA
UNIVERSITÀ DI ROMA

Linguistic Computing Laboratory
<http://lcl.uniroma1.it>

ERC Starting Grant MultiJEDI No. 259234

Course Objectives

Introduce the topic of Multilingual Semantic Processing, including:

- Task formulations and requisite resources
- Standard and state-of-the-art resources
- Applications for Multilingual Semantic Processing
- Open issues and future directions

Get you interested in Multilinguality

- How to extend your current work to the multiple languages
- How to improve performance with multilingual data

Target audience: we assume no prior knowledge of NLP or speaking multiple languages, so we target anyone with a genuine interest in this field

Tutorial Outline (1)

- Foundations in Semantic Processing
 - Basic concepts, terminology, and examples
 - Motivations for incorporating multilinguality
- Constructing Multilingual Semantic Resources
 - Methods for building new resources by combining heterogeneous resources in many languages
 - How multilingual representations solve current problems

Tutorial Outline (2)

- Identifying Concepts and Entities in Multilingual Text
 - How Entities and Concepts Differ
 - Methods for identify each in any language
- Multilingual Semantic Information Extraction
 - Moving from relating strings to relating things
- BabelNet in the LOD world
 - Lemon representation
 - Technical description
- Enriching Multilingual Semantic Resources through Gamification
 - How games address the need for enrichment
 - Designing games to solve the knowledge acquisition bottleneck

Part 1: Foundations

Understanding a simple phrase



Barack Obama peruses the internet.



Natural language is ambiguous

Yesterday, I saw an underground rock concert

Natural language is ambiguous

Yesterday, I saw an underground rock concert



or



Natural language is ambiguous

Underground rock concert

- a music event

Underground rock formation

- a stone structure

Natural language is highly ambiguous

Underground rock concert

- a music event

Underground rock formation

- a stone structure

Formation of an underground rock concert

- setup and planning for a music event

Natural language is highly ambiguous

Underground rock concert

- a music event

Underground rock formation

- a stone structure

Formation of an underground rock concert

- setup and planning for a music event

(?) A concert of underground rock formations

- (metaphoric) harmoniously arranged stone structures

Natural language is highly ambiguous

Underground rock concert

- a music event

Underground rock formation

- a stone structure

Formation of an underground rock concert

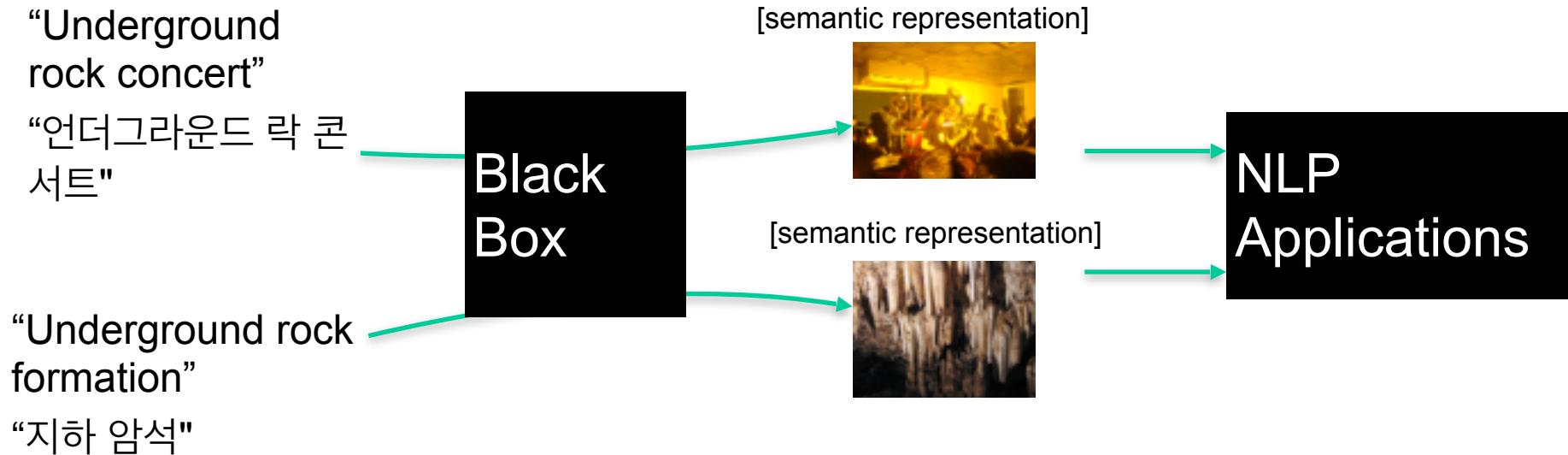
- setup and planning for a music event

(?) A concert of underground rock formations

- (metaphoric) harmoniously arranged stone structures

We need knowledge of a phrase's *semantics*

The Multilingual, Big-Picture Goal



The General Problem

POLYSEMY

- Most words have more than one possible meaning!
- Our job:
Model meaning from a computational perspective

Monosemous vs. Polysemous words

- **Monosemous words have only one meaning**
 - Examples:
 - *plant life*
 - *internet*
- **Polysemous words have more than one meaning**
 - Example: *bar*
 - “a room or establishment where alcoholic drinks are served”
 - “a counter where you can obtain food or drink”
 - “a rigid piece of metal or wood”
 - “musical notation for a repeating pattern of musical beats”

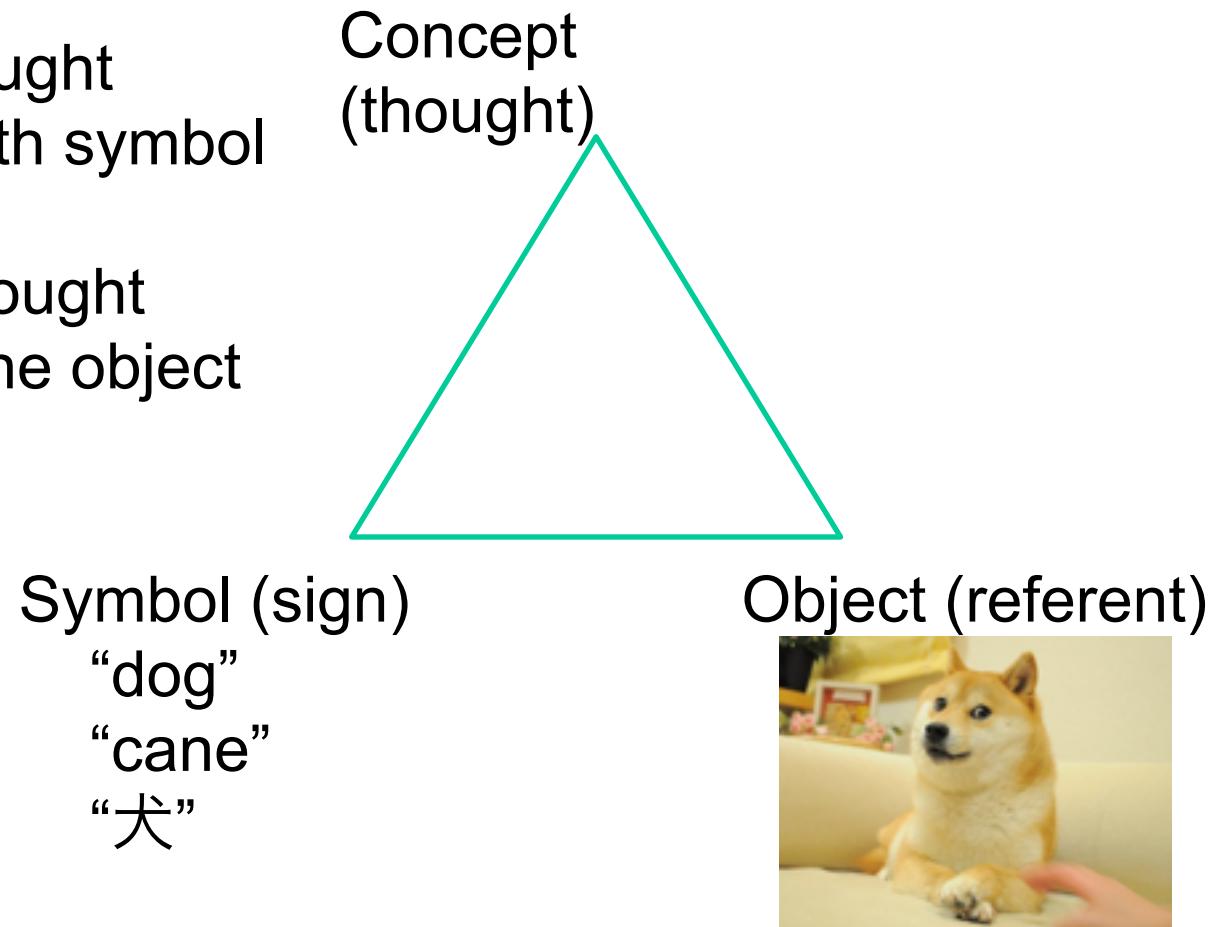
The Triangle of Meaning (Semiotic Triangle)

Writer:

- object evokes thought
- refers to object with symbol

Reader:

- symbol evokes thought
- refers symbol to the object



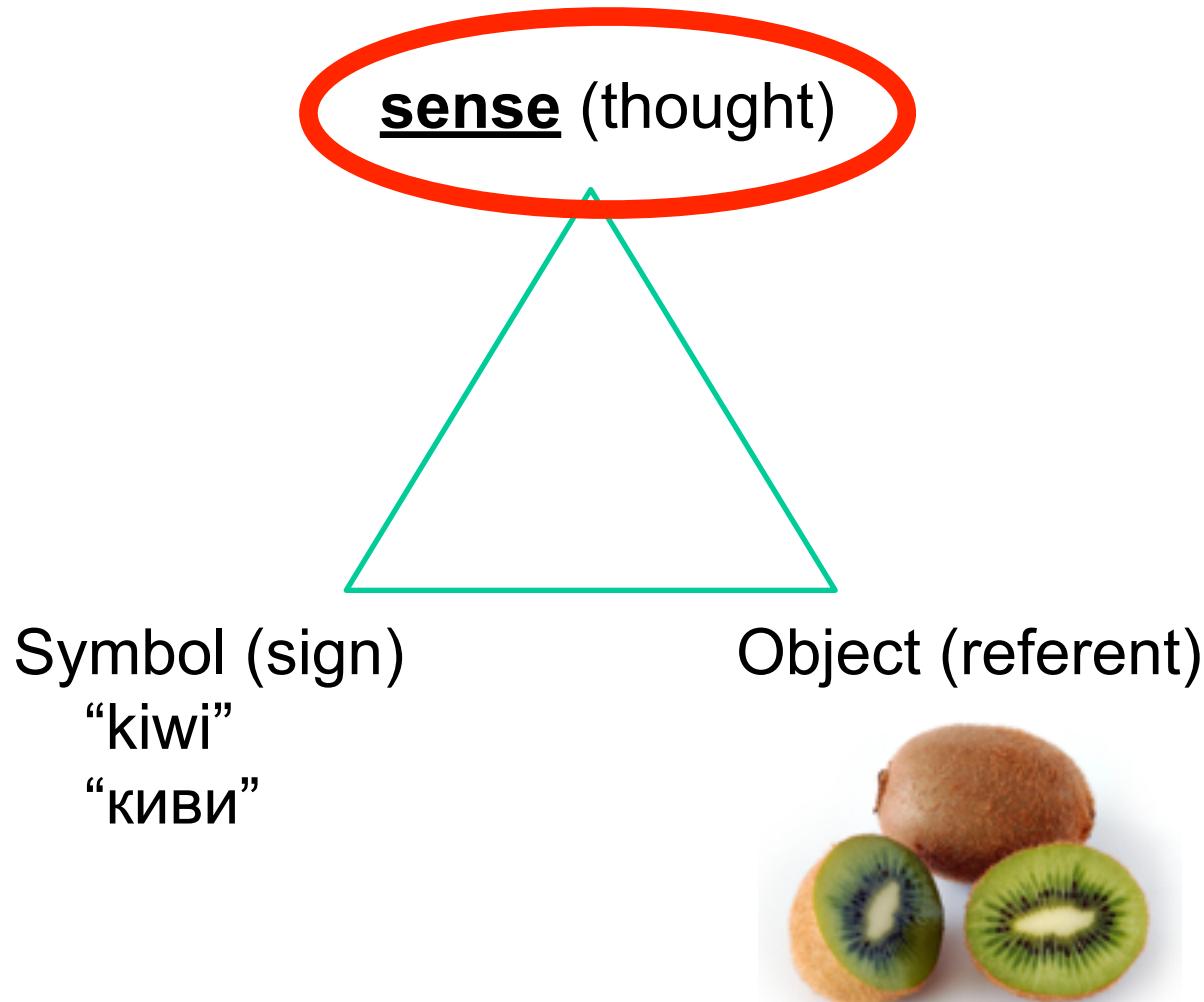
What is a word sense?

- A word sense is a commonly-accepted meaning of a word:
 - We are fond of fruit such as **kiwi**_{/fruit} and banana.
 - The **kiwi**_{/bird} is the national bird of New Zealand.
- How to represent word senses?
 - Can we enumerate the senses of a word?



- “Kiwi is my mother tongue, but I also speak all other English languages”

What is a word sense?



Word Senses

- The **bank** holds the mortgage on my home.
- The river overflowed its **banks** this year.
- He walked to the **bank** on the street corner.
- The treasures were buried in **banks** of dirt.



Word Senses: Homonymy

- The **bank**¹ holds the mortgage on my home.
- The river overflowed its **banks**² this year.
- He walked to the **bank**³ on the street corner.
- The treasures were buried in **banks**⁴ of dirt.



Homonymy: two senses share an orthographic form (e.g., bank), but are semantically and etymologically unrelated (different lemmas!)

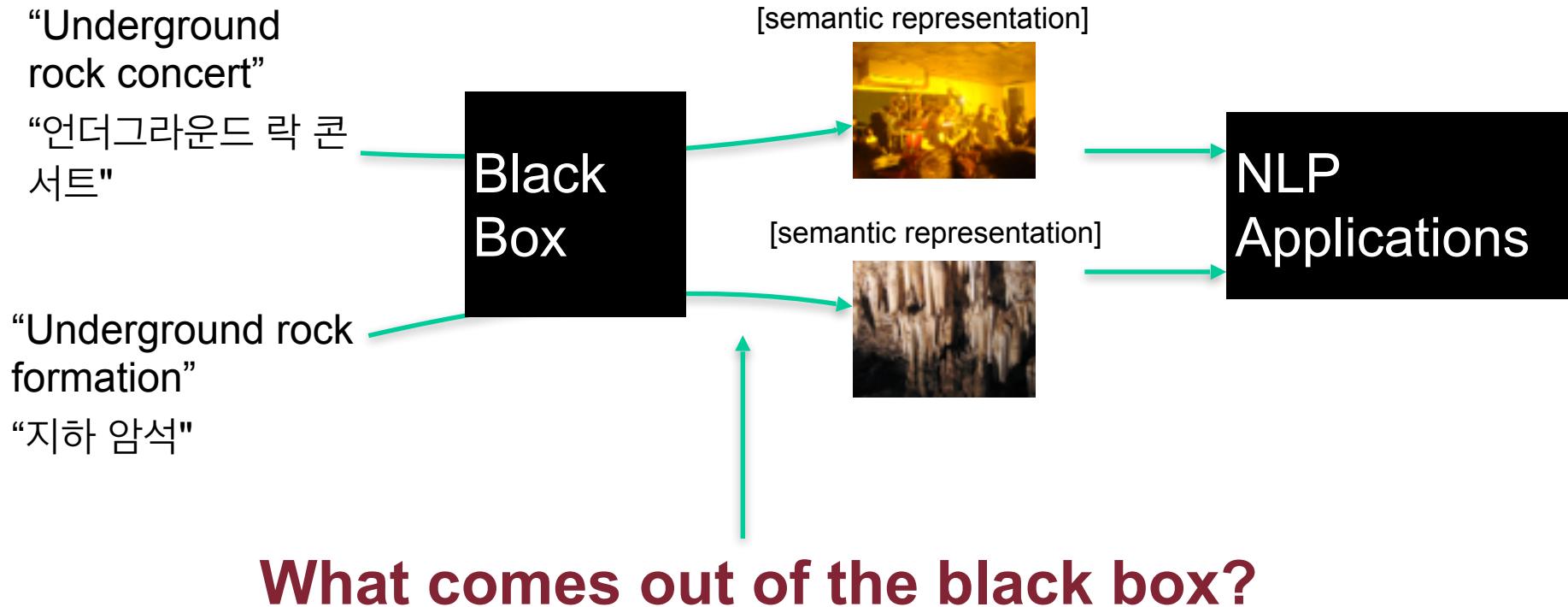
Word Senses: Polysemy

- The **bank**¹ holds the mortgage on my home.
- The river overflowed its **banks**² this year.
- He walked to the **bank**³ on the street corner.
- The treasures were buried in **banks**⁴ of dirt.



Polysemy: two senses are very close to each other semantically

How do we represent and encode semantics?



How do we represent and encode semantics?

- **Thesauri**
 - Groups words according to similar meaning
 - Relations between groups (e.g., narrower meanings)
 - Roget's Thesaurus (1911)
- **Machine Readable Dictionaries**
 - Enumerates all meanings of a word
 - Includes definitions, morphology, example usages, etc.
 - Oxford Dictionary of English, LDOCE, Collins, etc.
- **Computation Lexicons**
 - Repositories of structured knowledge about a word semantics and syntax
 - Include relations like hypernymy, meronymy, or entailment
 - WordNet

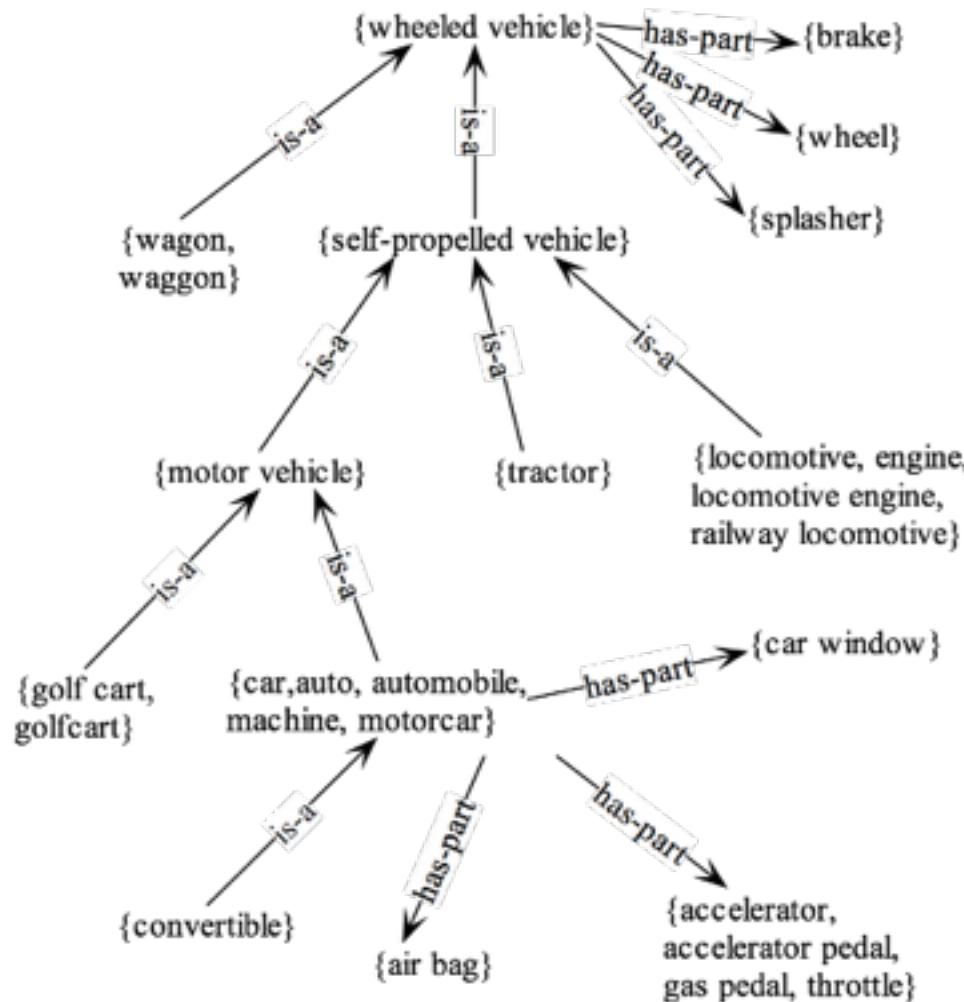
Senses and Relations in WordNet

- Each meaning is encoded as a **synset**, which is a collection of synonymous senses

$$\begin{aligned} Senses_{WN}(car_n) = & \{ \{ car_n^1, auto_n^1, automobile_n^1, machine_n^4, motorcar_n^1 \}, \\ & \{ car_n^2, rail car_n^1, rail way car_n^1, rail road car_n^1 \}, \\ & \{ cable car_n^1, car_n^3 \}, \\ & \{ car_n^4, gondola_n^3 \}, \\ & \{ car_n^5, elevator car_n^1 \} \}. \end{aligned}$$

- **Semantic relations** between synsets
 - Hypernymy (car_n^1 is-a motor vehicle $_n^1$)
 - Meronymy (car_n^1 has-a car door $_n^1$)
 - Entailment, similarity, attribute, etc.
- **Lexical relations** between word senses
 - Antonymy (good $_a^1$ antonym of bad $_a^1$)
 - Pertainymy (dental $_a^1$ pertains to tooth $_n^1$)
 - Nominalization (service $_n^2$ nominalizes serve $_v^4$)

WordNet as a semantic network



WordNets in other Languages

- EuroWordNet (Vossen, 1998)
- BalkaNet (Tufis et al., 2004)
- Multilingual Central Repository (Atserias et al., 2003)
- GermaNet (Hamp and Feldweg, 1997)
- SloWNet (Fišer and Sagot, 2008)
- Hungarian WN (Miháltz et al, 2008)
- Japanese WN (Isahara et al, 2008)
- ...
- Currently **73** unique WordNets

<http://globalwordnet.org/wordnets-in-the-world/>

An ideal resource for Multilingual Semantic Processing

- Capable of representing the meaning of a piece of text as word senses **in any language**
 - broad coverage of different senses, including language-specific senses
 - currently problematic for many language-specific WordNets
- Encodes semantic and syntactic relationships between the synsets
 - Highly beneficial for NLP applications
 - Encodes definitions and usages for synsets

Part 2: Building resources for multilingual semantic processing

Objective and motivation

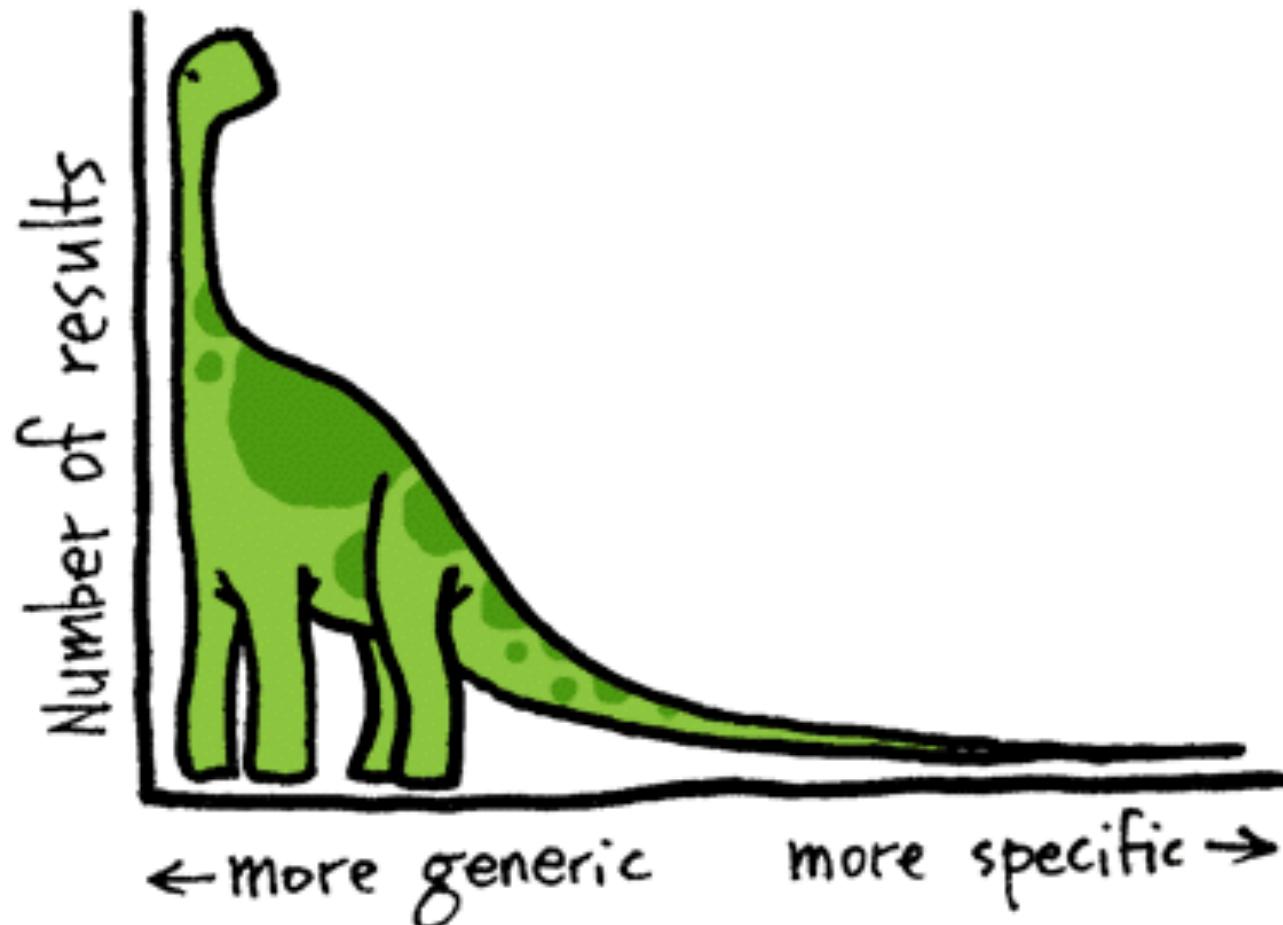
Goal:

- A large repository of knowledge in a multilingual setting

Motivations:

- A common ground for language technologies that provides:
 - Multilinguality
 - Encyclopedic knowledge
 - Lexicographic knowledge
 - Semantic relations
 - ...

Taming the long tail...



Core Challenges

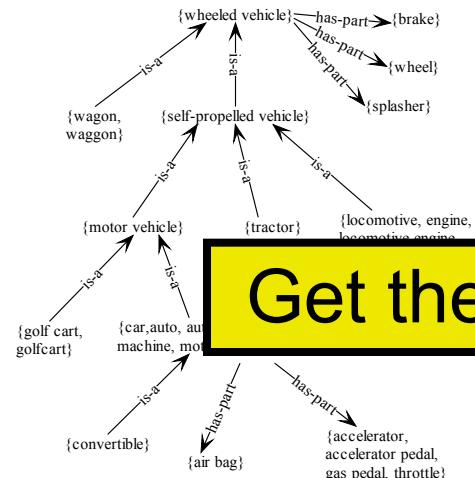
1. Integrating and unifying different resources;
2. Managing many different languages;
3. Extending semantic relations within the single resources and between concepts contained in different resources;
4. Maintaining high accuracy.

Goal: Creating a Multilingual Semantic Network

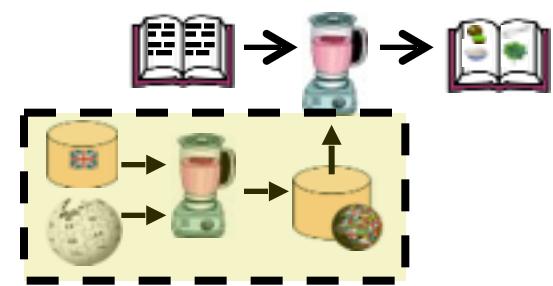
Start from two large **complementary** resources:

WordNet: full-fledged taxonomy

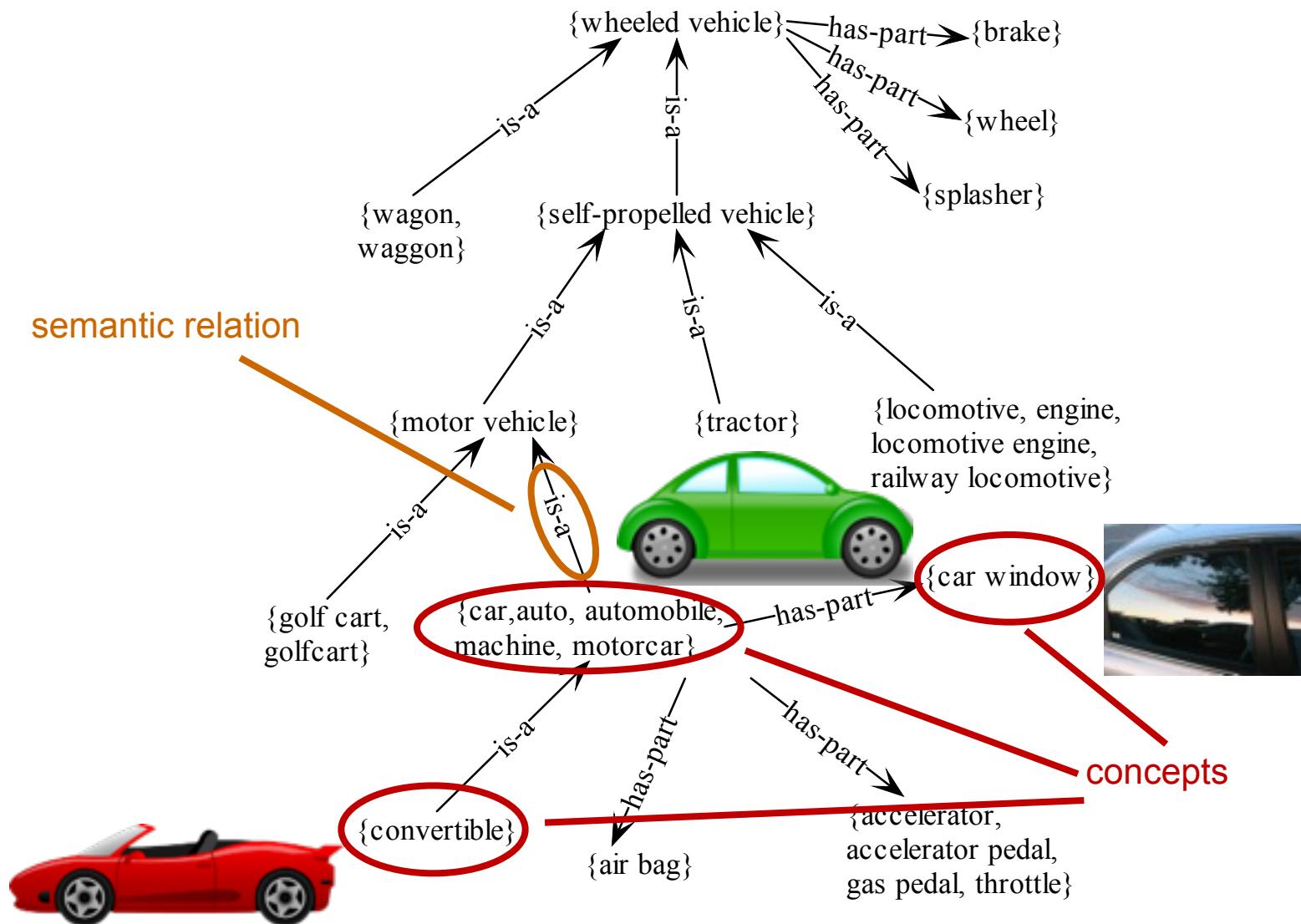
Wikipedia: multilingual and continuously updated



Get the best from both worlds



WordNet [Miller et al., 1990; Fellbaum, 1998]



Wikipedia [The Web Community, 2001-today]

Automobile

From Wikipedia, the free encyclopedia
(Redirected from Car)

For the magazine, see [Automobile Magazine](#).

"car" redirects here. For other uses, see [Car \(disambiguation\)](#).

An automobile, autocar, motor car or car is a wheeled [motor vehicle](#) used for transporting passengers, which also carries its own engine or motor. Most definitions of the term specify that automobiles are designed to run primarily on roads, to have seating for one to eight people, to typically have four wheels, and to be constructed principally for the transport of people rather than goods.^[3]

The term [motorcar](#) has also been used in the context of electrified rail systems to denote a car which functions as a small locomotive but also provides space for passengers and baggage. These locomotive cars were often used on suburban routes by both interurban and intercity railroad systems.^[4]

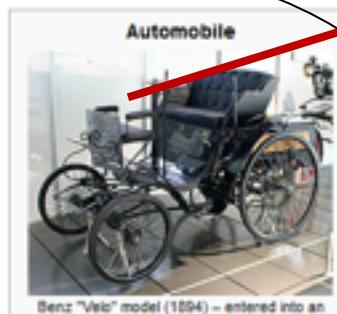
It was estimated in 2010 that the number of automobiles had risen to over 1 billion vehicles, with 500 million reached in 1986.^[5] The numbers are increasing rapidly, especially in [China](#) and [India](#).^[6]

Motor vehicle

From Wikipedia, the free encyclopedia

A [motor vehicle](#) or [road vehicle](#) is a self-propelled wheeled vehicle that does not operate on rails, such as [trains](#) or [trolleys](#). The vehicle propulsion is provided by an [engine](#) or [motor](#), usually by an [internal combustion engine](#), or an [electric motor](#), or some combination of the two, such as [hybrid electric vehicles](#) and [plug-in hybrids](#). For legal purposes motor vehicles are often identified within a number of vehicle classes including [automobiles](#) or [cars](#), [buses](#), [motorcycles](#), [motorized bicycles](#), [off highway vehicles](#), [light trucks](#) or [light duty trucks](#), and [trucks](#) or [lorries](#). These classifications vary according to the legal codes of each country. [ISO 3833:1977](#) is the standard for road vehicles types, terms and definitions.^[1]

As of 2010 there were more than one billion motor vehicles in use in the world excluding [off-road vehicles](#) and [heavy construction equipment](#).^{[2][3][4]} Global vehicle ownership [per capita](#) in 2010 was 148 vehicles in operation per 1000 people.^[4] The United States has the largest fleet of motor vehicles in the world, with 239.8 million by 2010. Vehicle



Benz "Velo" model (1894) – entered into an early automobile race as a motorcycle.^{[1][2]}

Classification Vehicle
Industry Various



The United States has the world's largest motor vehicle registered fleet, with almost 250 million vehicles.

(unspecified) semantic relation

concepts

Passenger

From Wikipedia, the free encyclopedia

This article is about passengers in commercial transportation; for other uses see [Passenger \(disambiguation\)](#).

A [passenger](#) is a person who travels in a [vehicle](#) but bears little or no responsibility for the tasks required for that vehicle to arrive at its destination or otherwise operate the vehicle.

Passengers are people who ride on [buses](#), [passenger trains](#), [airliners](#), [ships](#), [ferryboats](#), and other methods of transportation.

crew members (if any), as well as the driver or pilot of the vehicle, are considered to be passengers. For example, a flight attendant would not be considered a "passenger" while on duty, but an idling in a [company car](#) being driven by another person would be a passenger, even if the car was being driven on company



Travel

From Wikipedia, the free encyclopedia
(Redirected from [Traveling](#))

For other uses, see [Travel \(disambiguation\)](#).

Travel is the movement of [people](#) or [objects](#) (such as [airplanes](#), [boats](#), [trains](#) and other [conveyances](#)) between relatively distant geographical locations.^{[1][2]}

Contents

- 1 Etymology
- 2 Purpose and motivation
- 3 Travel safety
- 4 See also
- 5 References
- 6 External links

Etymology

The term "travel" originates from the Old French word [travel](#).^[3] The term also covers all the activities performed during a [travel \(movement\)](#).^[4]



An example of mapping



Balloon (aircraft)

From Wikipedia, the free encyclopedia

Not to be confused with [Airship](#) or [Blimp](#).



This article needs additional citations for verification. Please help [improve this article](#) by adding citations to reliable sources. Unsourced material may be challenged and removed. (April 2008)

A balloon is a type of [aerostat](#) that remains aloft due to its [buoyancy](#). A balloon travels by moving with the wind. It is distinct from an [airship](#), which is a buoyant aircraft that can be propelled through the air in a controlled manner.

The "basket" or capsule that is suspended by cables beneath a balloon and carries people, animals, or automatic equipment (including cameras and telescopes, and flight-control mechanisms) may also be called the [gondola](#).



Contents [view]

- 1 Types
- 2 History
- 3 As flying machines
- 4 Military use
 - 4.1 American Civil War
 - 4.2 After the American Civil War
- 5 Records
- 6 In space
- 7 Sports



Creation of the Wikipedia disambiguation

 [Log in](#) [Create account](#)

Article [Talk](#) [Read](#) [Edit](#) [View history](#) 

Balloon (aircraft)

From Wikipedia, the free encyclopedia

Not to be confused with [Airship](#) or [Blimp](#).

 This article needs additional [citations for verification](#). Please help [improve this article](#) by adding citations to [reliable sources](#). Unsourced material may be [challenged](#) and removed. (April 2008)

A balloon is a type of aerostat that remains aloft due to its [buoyancy](#). A balloon travels by moving with the wind. It is distinct from an [airship](#), which is a buoyant aircraft that can be propelled through the air in a controlled manner.

The "basket" or capsule that is suspended by cables beneath a balloon and carries people, animals, or automatic equipment (including cameras and telescopes, and flight-control mechanisms) may also be called the [gondola](#).

Contents [hide]

- [1 Types](#)
- [2 History](#)
- [3 As flying machines](#)
- [4 Military use](#)
 - [4.1 American Civil War](#)
 - [4.2 After the American Civil War](#)
- [5 Records](#)



Categories: [Balloons \(aircraft\)](#) | [Ballooning](#) | [Airship technology](#) | [Hydrogen technologies](#) | [Aeronautics](#)

$\text{ctx}(\text{Balloon (aircraft)}) = \{ \}$

Creation of the Wikipedia disambiguation

 **WIKIPEDIA**
The Free Encyclopedia

Article Talk sense label Read Edit View history Search

Balloon (aircraft)

From Wikipedia, the free encyclopedia

Not to be confused with [Airship](#) or [Blimp](#).

 This article needs additional [citations for verification](#). Please help [improve this article](#) by adding citations to [reliable sources](#). Unsourced material may be [challenged](#) and removed. (April 2008)

A balloon is a type of aerostat that remains aloft due to its [buoyancy](#). A balloon travels by moving with the wind. It is distinct from an [airship](#), which is a buoyant aircraft that can be propelled through the air in a controlled manner.

The "basket" or capsule that is suspended by cables beneath a balloon and carries people, animals, or automatic equipment (including cameras and telescopes, and flight-control mechanisms) may also be called the [gondola](#).

Contents [hide]

- [1 Types](#)
- [2 History](#)
- [3 As flying machines](#)
- [4 Military use](#)
 - [4.1 American Civil War](#)
 - [4.2 After the American Civil War](#)
- [5 Records](#)



Categories: [Balloons \(aircraft\)](#) | [Ballooning](#) | [Airship technology](#) | [Hydrogen technologies](#) | [Aeronautics](#)

$\text{ctx}(\text{Balloon (aircraft)}) = \{ \text{aircraft} \}$

Creation of the Wikipedia disambiguation

[Log in](#) [Create account](#)



WIKIPEDIA
The Free Encyclopedia

Article [Talk](#)

Read [Edit](#) [View history](#)

Search



Balloon (aircraft)

From Wikipedia, the free encyclopedia

Not to be confused with [Air](#)



This article needs additional citations for verification. Please help improve this article by adding reliable sources. Unverified material may be challenged and removed. (April 2008)

hyperlinks

A balloon is a type of aerostat that remains aloft due to its buoyancy. A balloon travels by moving with the wind. It is distinct from an airship, which is a buoyant aircraft that can be propelled through the air in a controlled manner.

The "basket" or capsule that is suspended by cables beneath a balloon and carries people, animals, or automatic equipment (including cameras and telescopes, and flight-control mechanisms) may also be called the gondola.

Contents [hide]

- [1 Types](#)
- [2 History](#)
- [3 As flying machines](#)
- [4 Military use](#)
 - [4.1 American Civil War](#)
 - [4.2 After the American Civil War](#)
- [5 Records](#)



Categories: [Balloons \(aircraft\)](#) | [Ballooning](#) | [Airship technology](#) | [Hydrogen technologies](#) | [Aeronautics](#)

$\text{ctx}(\text{Balloon (aircraft)}) = \{ \text{aircraft}, \text{aerostat}, \text{buoyancy}, \text{airship}, \dots, \text{gondola} \}$

Creation of the Wikipedia disambiguation

Log in Create account

Article Talk Read Edit View history Search

Balloon (aircraft)

From Wikipedia, the free encyclopedia

Not to be confused with [Airship](#) or [Blimp](#).

This article needs additional [citations for verification](#). Please help [improve this article](#) by adding citations to [reliable sources](#). Unsourced material may be [challenged](#) and removed. (April 2008)

A balloon is a type of aerostat that remains aloft due to its [buoyancy](#). A balloon travels by moving with the wind. It is distinct from an [airship](#), which is a buoyant aircraft that can be propelled through the air in a controlled manner.

The "basket" or capsule that is suspended by cables beneath a balloon and carries people, animals, or automatic equipment (including cameras and telescopes, and flight-control mechanisms) may also be called the [gondola](#).

Contents [hide]

- 1 Types
- 2 History
- 3 As flying machine
- 4 Military use
 - 4.1 American Civil War
- 5 Records

categories



Categories: [Balloons \(aircraft\)](#) | [Ballooning](#) | [Airship technology](#) | [Hydrogen technologies](#) | [Aeronautics](#)

$\text{ctx}(\text{Balloon (aircraft)}) = \{ \text{aircraft}, \text{aerostat}, \text{buoyancy}, \text{airship}, \dots, \text{gondola}, \text{ballooning}, \text{hydrogen}, \text{aeronautics} \}$

Building BabelNet: Mapping Wikipedia to WordNet

$$\begin{aligned}\mu(w) &= \operatorname{argmax}_{s \in Senses_{WN}(w)} p(s|w) = \operatorname{argmax}_s \frac{p(s, w)}{p(w)} \\ &= \operatorname{argmax}_s p(s, w)\end{aligned}$$

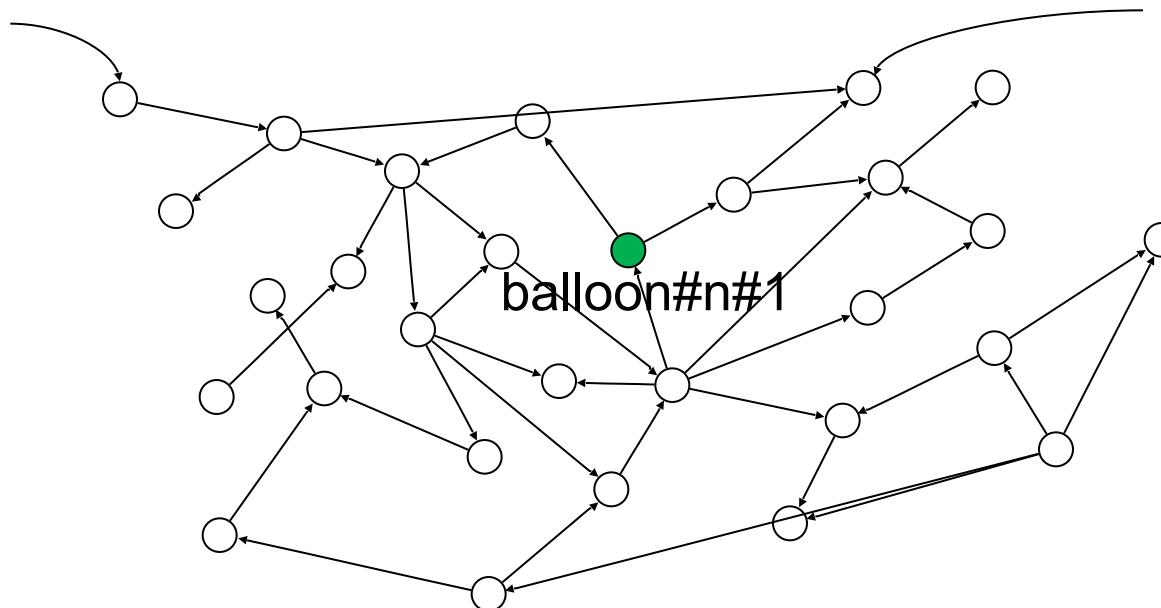
Given a Wikipage w and its disambiguation context $\text{ctx}(w)$:
For each WordNet sense s of w , calculate $\text{score}(s, w)$ as follows:

$$\text{score}(s, w) = \sum_{cw \in \text{Ctx}(w)} \sum_{s' \in Senses_{WN}(cw)} \sum_{p \in paths_{WN}(s, s')} e^{-(length(p)-1)}$$

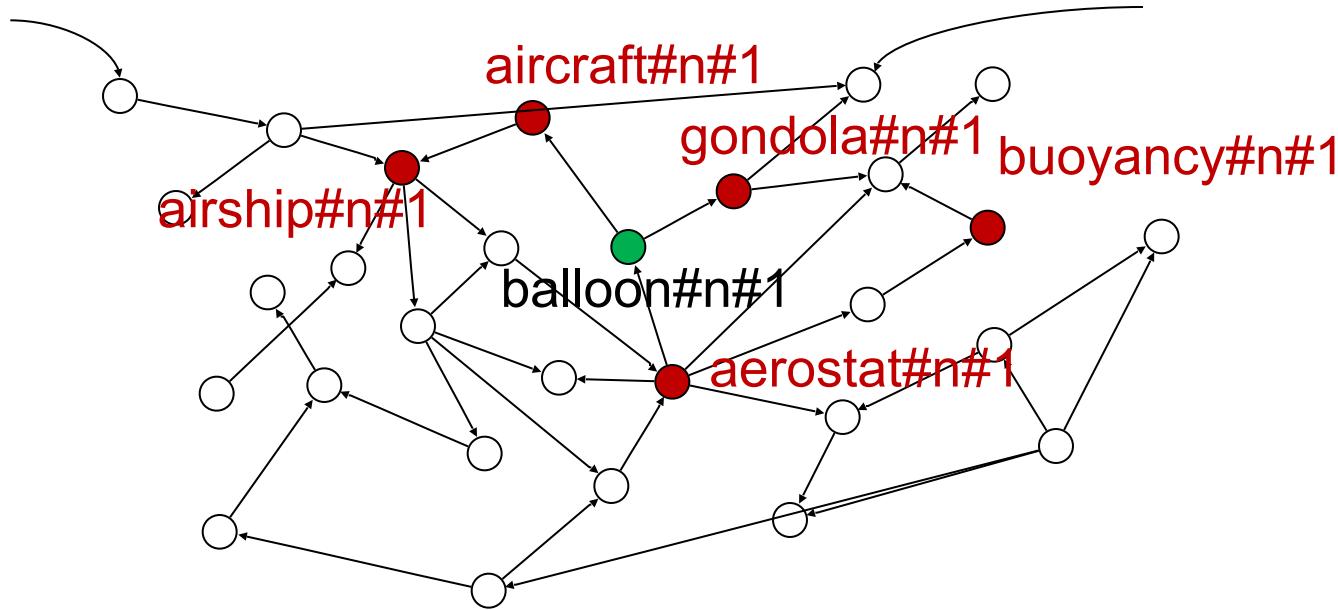


The Wikipedia page context in the WordNet graph

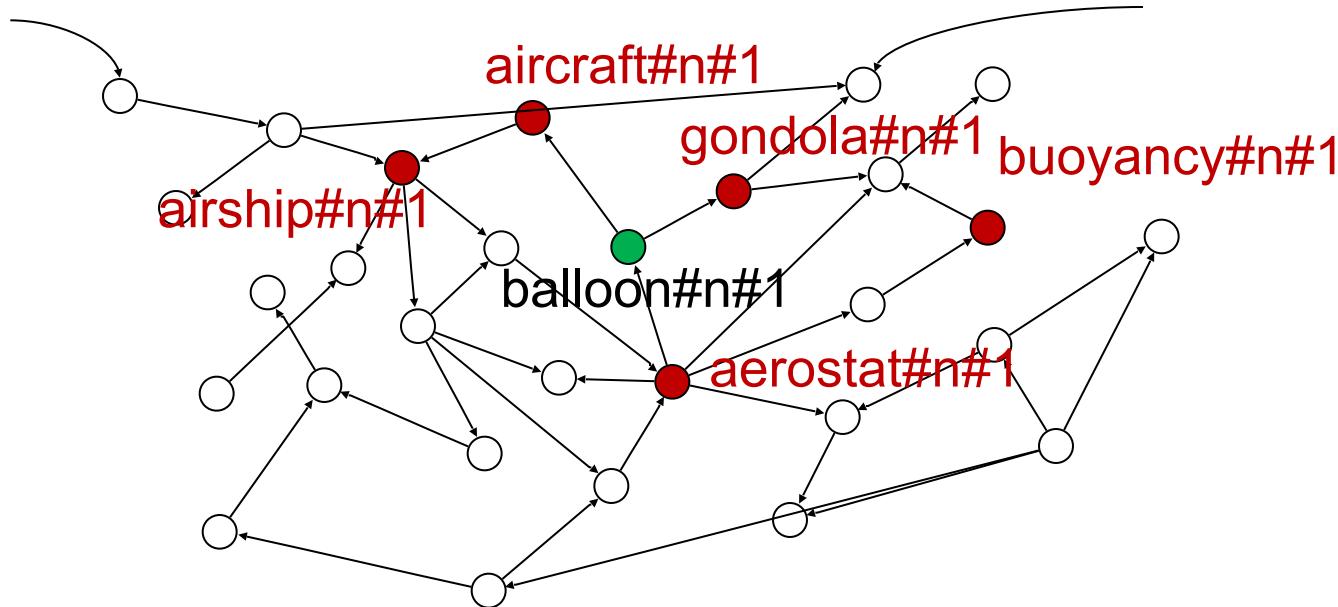
$\text{ctx}(\text{Balloon (aircraft)}) = \{ \text{aircraft, aerostat, buoyancy, airship, ..., gondola} \}$



The Wikipedia page context in the WordNet graph



The Wikipedia page context in the WordNet graph



balloon#n#1 -> aircraft#n#1

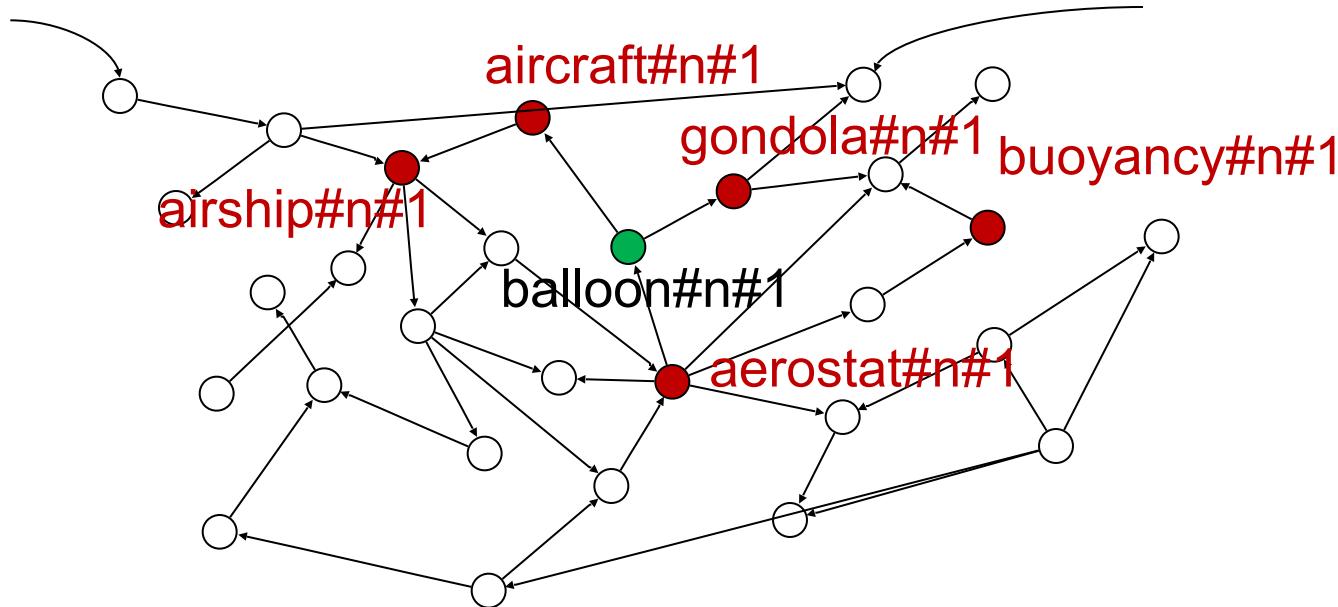
balloon#n#1 -> aircraft#n#1 -> airship#n#1

balloon#n#1 -> gondola#n#1

balloon#n#1 -> gondola#n#1 -> flight#n#1 -> buoyancy#n#1

balloon#n#1 -> aerostat#n#1

The Wikipedia page context in the WordNet graph



balloon#n#1 -> aircraft#n#1

balloon#n#1 -> aircraft#n#1 -> airship#n#1

balloon#n#1 -> gondola#n#1

balloon#n#1 -> gondola#n#1 -> flight#n#1 -> buoyancy#n#1

balloon#n#1 -> aerostat#n#1

→ 0.35

BabelNet integrates the best of both worlds

WordNet

balloon

S: (n) balloon (large tough nonrigid bag filled with gas or heated air)
S: (n) balloon (small thin inflatable rubber bag with narrow neck)

Wikipedia

Speech balloon

From Wikipedia, the free encyclopedia

Speech balloons (also speech bubbles, dialogue balloons or word balloons) are a graphic convention used most commonly in comic books, comic strips and cartoons to allow words (and much less often, pictures) to be understood as representing the speech or thoughts of a given character in the comic. There is often a formal distinction between the balloon that indicates thoughts and the one that indicates words spoken aloud; the former conveys subjective thoughts while the latter conveys objective thoughts.



Balloon (typeface)

From Wikipedia, the free encyclopedia

This article does not cite any references or sources. Please help improve this article by adding citations to reliable sources. Unsourced material may be challenged and removed. (June 2012)

Contents [hide]

- 1 History
- 2 Popular forms
 - 2.1 Speech bubbles
 - 2.2 Thought bubbles
 - 2.3 Other forms

Balloon is a brush script commonly used for signage or display purposes. It was designed in 1939 by Max R. Kauffmann, for American Type Founders, in response to Howard Allen Trott's cartoon, cut for Bauer type. Foundry in 1938. It had no lowercase letters and was cast in Light, Bold, and Extra.

Balloon
BALLOON

Category Script

Designer(s) Max R. Kauffmann

Foundry ATF

THE QUICK BROWN FOX JUMPS OVER THE LAZY DOG.

ANSWERSTON TYPE FOUNDERS LTD LTD

Sample



WIKIPEDIA

The free encyclopedia

Article Tab Read Edit Search

Wiki Loves Monuments

Photograph a monument, help Wikipedia and win!

Balloon

Jump to navigation

Search

Recent changes

Upload files

Special pages

Help

Community portal

Recent changes

Random page

Log in

Help

Print page

Language

EN

DE

ES

FR

IT

JA

BR

ES

PT

CA

AR

FA

HE

IL

SV

DA

NO

NN

MS

CS

SK

SR

HR

ME

AL

MT

XH

XN

XO

XZ

XU

XW

XG

XK

XQ

XV

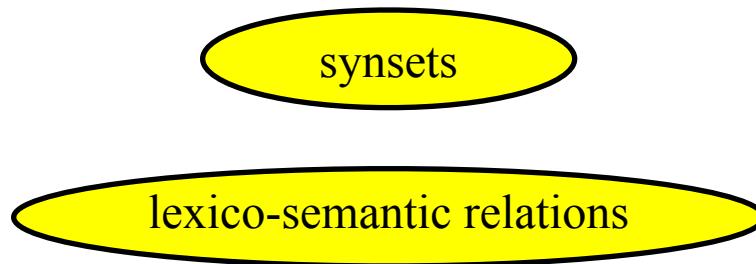
XN

XW

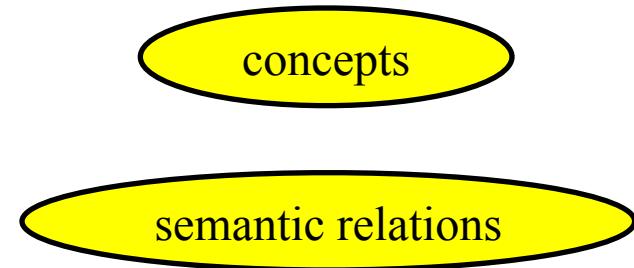
BabelNet: concepts and semantic relations (1)

Concepts and relations in BabelNet are harvested from
WordNet and **Wikipedia**:

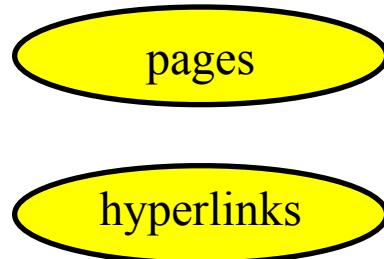
WordNet:



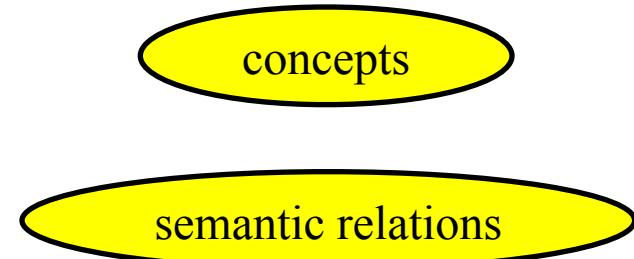
BabelNet:



Wikipedia:



BabelNet:



BabelNet: concepts and semantic relations (2)

We encode knowledge as a labeled directed graph:

Each vertex is a Babel synset



Each edge is a semantic relation between synsets:

is-a (balloon is-a aircraft)

part-of (gasbag part-of balloon)

instance-of (Einstein instance-of physicist)

...

unspecified/relatedness (balloon related-to flight)

Building BabelNet: Translating Babel synsets

1. Exploiting Wikipedia interlanguage links

The diagram illustrates the exploitation of Wikipedia interlanguage links. Three language labels are shown in yellow boxes: "Balloon" (top), "globo aerostàtico" (middle), and "pallone aerostatico" (bottom). Arrows from these labels point to a screenshot of a Wikipedia page about balloons. The screenshot shows the English Wikipedia page for "Balloon". The sidebar on the left lists various language links, including English, French, German, Spanish, and Italian. The main content area contains a table of contents with sections like "Types", "History", and "Gas balloons". A photograph of a hot air balloon is visible on the right side of the page.

Contents [edit]

1 Types

2 History

3 As flying machines

4 Military use

4.1 American Civil War

4.2 After the American Civil War

5 Records

6 In space

7 Sports

8 See also

9 References

10 External links

Types [edit]

There are three main types of balloons:

- hot air balloons obtain their buoyancy by heating the air inside the balloon. They are the most common type of balloon aircraft. "Hot air balloon" is sometimes used incorrectly to denote any balloon that carries people.
- gas balloons are inflated with a gas of lower molecular weight than the ambient atmosphere. Most gas balloons operate with the internal pressure of the gas the same as the pressure of the surrounding atmosphere. There is a type of gas balloon, called a superpressure balloon, that can operate with the lifting gas at pressure that exceeds the pressure of the surrounding air, with the objective of limiting or eliminating the loss of gas from day time heating. Gas balloons are filled with gases such as:
 - hydrogen – not widely used for aircraft since the Hindenburg disaster because of high flammability (except for some sport balloons as well as nearly all unmanned scientific and weather balloons)
 - helium – the gas used today for all airships and most manned balloons.
 - ammonia – used infrequently due to its caustic qualities and limited lift.
 - coal gas – used in the early days of ballooning; it is highly flammable.
 - methane – used as a lower cost lifting gas, but offering less lift than helium or hydrogen [1]
- Rozhniat balloons use both heated and unheated lifting gases. The most common modern use of this type of balloon is for long-distance record flights such as the recent circumnavigations.

History [edit]

Main article: History of ballooning

Building BabelNet: Translating Babel synsets

2. Filling the lexical translation gaps using a Machine Translation system to translate the English lexicalizations of a concept

On August 27, 1783 in Paris, Franklin witnessed the world's first hydrogen [[Balloon (aircraft)|balloon]] flight.

Google Translate

Le 27 Août, 1783 à Paris, Franklin vu le premier vol en **ballon** d'hydrogène.

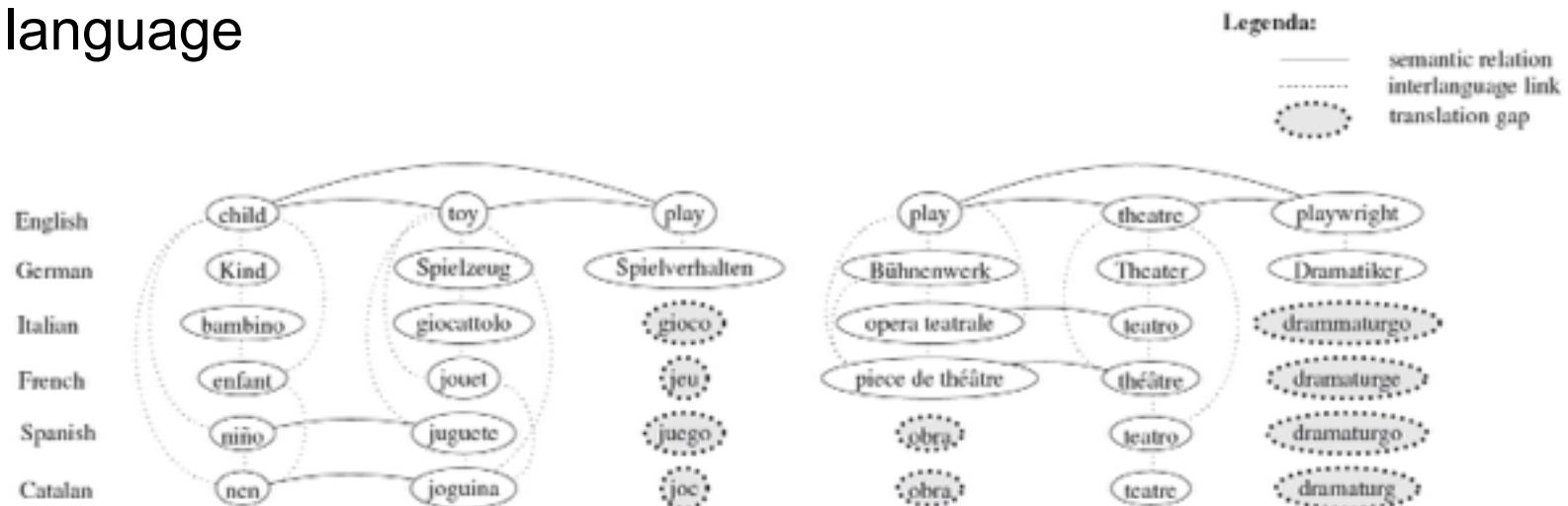
Building BabelNet: Translating Babel synsets

2. Filling the lexical translation gaps using a Machine Translation system to translate the English lexicalizations of a concept

For each word sense s , we translate:

- sentences from SemCor (a corpus annotated with WordNet senses) which contain s
- sentences from Wikipedia linked to the Wikipage of s

The most frequent translation of s is selected for each target language



The most frequent translation of a word in a given meaning

left context	term	right context
	<i>wikification</i>	<i>may refer to: the...</i>
geoinformatics services'	<i>wikification</i>	<i>of GIS by the masses'</i>
<i>the process may be called</i>	<i>wikification</i>	<i>(as in ...</i>
<i>which is then called "</i>	<i>wikification</i>	<i>and to the related problem</i>
<i>reason needs copyediting,</i>	<i>wikification</i>	<i>, reduction of POV, work on</i>
<i>huge amount of cleanup,</i>	<i>wikification</i>	<i>, etc. Version of 12 Nov</i>

The most frequent translation of a word in a given meaning

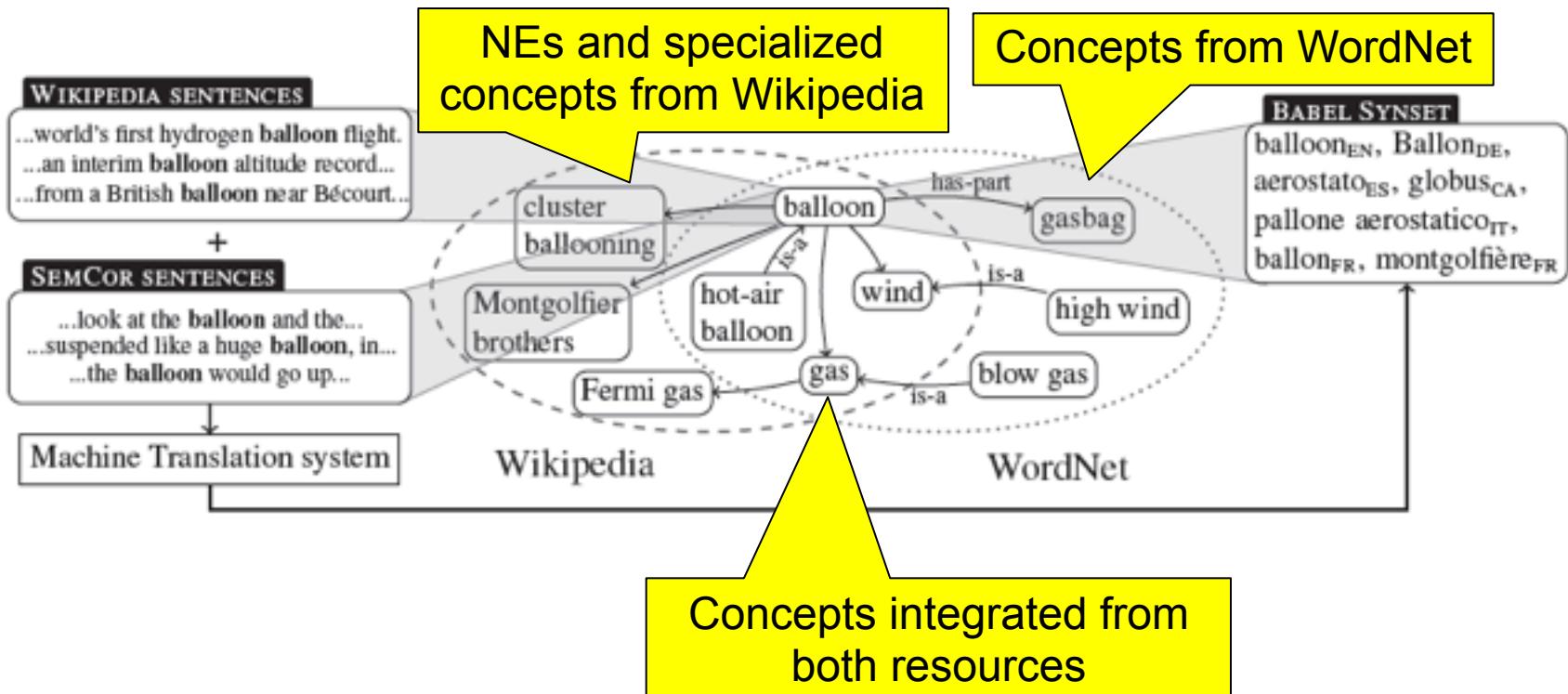
left context	term	right context
	<i>wikificazion</i>	<i>potrebbe riferirsi a: il...</i>
<i>servizi geoinformatici' e '</i>	<i>wikification</i>	<i>di GIS dalle masse'</i>
<i>il processo chiamato</i>	<i>wikificazion</i>	<i>(come in ...</i>
<i>che è quindi chiamato</i>	<i>wikificazion</i>	<i>e al problema correlato...</i>
<i>ragione richiede</i>	<i>wikification</i>	<i>, riduzione di POV, lavoro su</i>
<i>grandi quantità di pulizia,</i>	<i>wikificazion</i>	<i>, ecc. Versione del 12 Novembre</i>

The most frequent translation of a word in a given meaning

left context	term	right context
	<i>wikificazion</i>	<i>potrebbe riferirsi a: il...</i>
<i>servizi geoinformatici' e '</i>	<i>wikification</i>	<i>di GIS dalle masse'</i>
<i>il processo chiamato</i>	<i>wikificazion</i>	<i>(come in ...</i>
<i>che è quindi chiamato</i>	<i>wikificazion</i>	<i>e al problema correlato...</i>
<i>ragione richiede</i>	<i>wikification</i>	<i>, riduzione di POV, lavoro su</i>
<i>grandi quantità di pulizia,</i>	<i>wikificazion</i>	<i>, ecc. Versione del 12 Novembre</i>

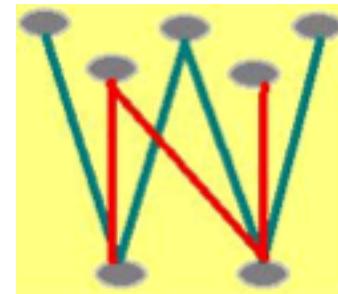
BabelNet [Navigli and Ponzetto, AIJ 2012]

A wide-coverage multilingual semantic network including both encyclopedic (from Wikipedia) and lexicographic (from WordNet) entries





WIKIPEDIA
The Free Encyclopedia



WordNet

Is that all ?

Open Multilingual WordNet

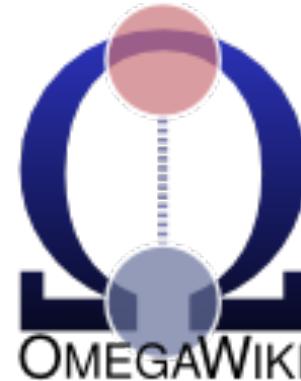
[Bond and Foster, 2013]

- <http://compling.hss.ntu.edu.sg/omw/>
- 22 languages
- Mappings to the Princeton WordNet synsets
- More than 600 thousands lexicalizations

Francis Bond and Kyonghee Paik. 2012. A survey of wordnets and their licenses. In Proc. of GWC 2012

Francis Bond and Ryan Foster. 2013. Linking and extending an open multilingual wordnet. In Proc. of ACL

OmegaWiki (<http://www.omegawiki.org>)



- 50 languages
- More than 250 thousands synsets
- Automatically mapped to BabelNet

How can we map together heterogeneous resources?

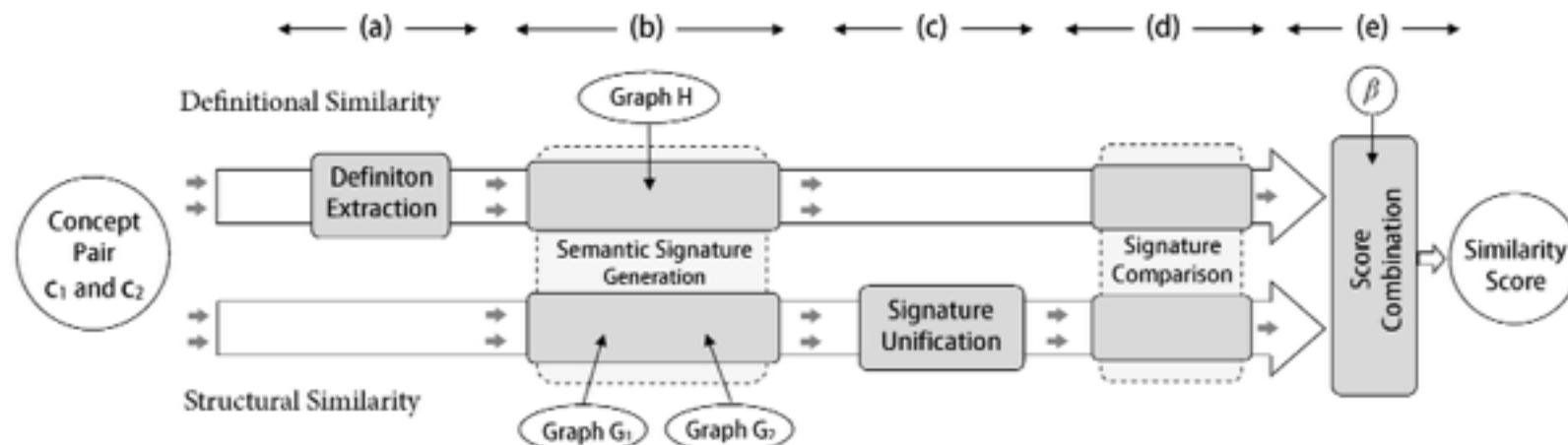
Remember, we are talking about semantics and so:

Similarity Measures!

Cross-resource Concept Alignment [Pilehvar and Navigli, 2014]

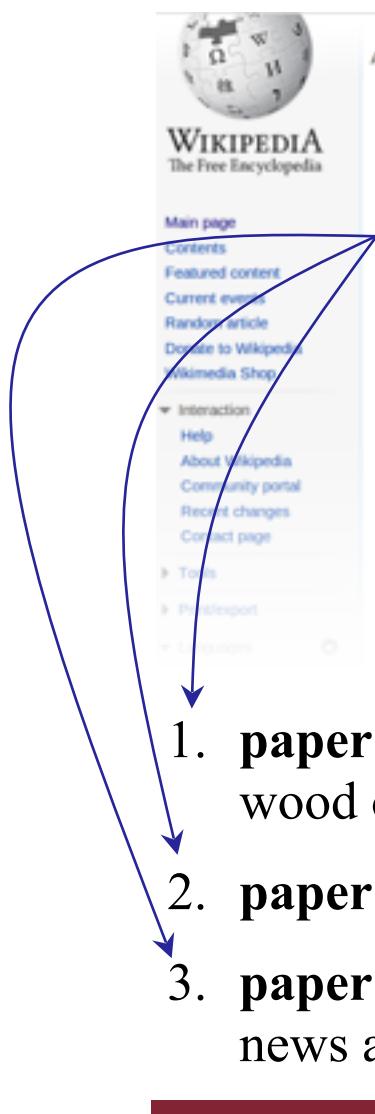
Align two concepts if their *similarity* is greater than a given threshold;

similarity is computed as linear combination of Definitional and Structural similarities.



Mohammad Taher Pilehvar and Roberto Navigli. 2014. A Robust Approach to Aligning Heterogeneous Lexical Resources. In Proc. of ACL

Definitional similarity (WP-WN)



The screenshot shows the Wikipedia article for "Paper". The sidebar on the left contains links like Main page, Contents, Featured content, Current events, Random article, Donate to Wikipedia, Wikimedia Shop, Interaction (Help, About Wikipedia, Community portal, Recent changes, Contact page), Tools, and Participate. The main content area has tabs Article (selected) and Talk. It includes a search bar and links for Read, View source, and View history. The article text starts with: "Paper is a thin material produced by pressing together moist fibers, typically cellulose pulp derived from wood, rags or grasses, and drying them into flexible sheets." This sentence is highlighted with a yellow box. Below it, there is more text about the versatility of paper and its history of development in China. To the right, there is a photograph showing different types of paper, labeled "Different typologies of paper". A note at the bottom right says "Wikimedia Commons has media related to Paper".

1. **paper** -- a material made of cellulose pulp derived mainly from wood or rags or certain grasses.
2. **paper** -- a n essay (especially one written as an assignment).
3. **paper** -- a daily or weekly publication on folded sheets; contains news and articles and advertisements.

Definitional similarity (WT-WN)

The screenshot shows the Wikipedia page for the noun 'paper'. The page includes a sidebar with language links (Indonesian, etc.) and a main content area with sections for 'Noun [edit]', 'Translations [edit]', and 'Derived terms [edit]'. The 'Noun' section contains a list of 11 definitions, with the first one highlighted in yellow. A blue arrow points from the top left towards this highlighted text. Another blue arrow points from the bottom left towards the third definition in the list.

Noun [edit]

paper (countable and uncountable, plural [papers](#))

1. A sheet material used for writing on or printing on (or as a non-waterproof container), usually made by draining cellulose fibres from a suspension in water.
[edit] [history] ↗
2. A newspaper or anything used as such (such as a newsletter or listing magazine).
[edit] [history] ↗
3. (uncountable) Wallpaper.
4. (uncountable) Wrapping paper.
5. A written document, generally shorter than a book (white paper, term paper) — in particular one written for the Government.
6. A written document that reports scientific or academic research and is usually subjected to peer review before publication in a scientific journal or in the proceedings of a scientific or academic meeting (such as a conference, a workshop or a symposium).
7. A scholastic essay.
8. (slang) Money.
9. (New Zealand) A university course.
10. A paper packet containing a quantity of items.
a [paper](#) of pins, tacks, opium, &c.
11. A medicinal preparation spread upon paper, intended for external application.
[cantharides paper](#)

Derived terms [edit]

Translations [edit]

material for writing on
newspaper — see [newspaper](#)
wallpaper — see [wallpaper](#)

1. **paper** -- a material made of cellulose pulp derived mainly from wood or rags or certain grasses.
2. **paper** -- a n essay (especially one written as an assignment).
3. **paper** -- a daily or weekly publication on folded sheets; contains news and articles and advertisements.

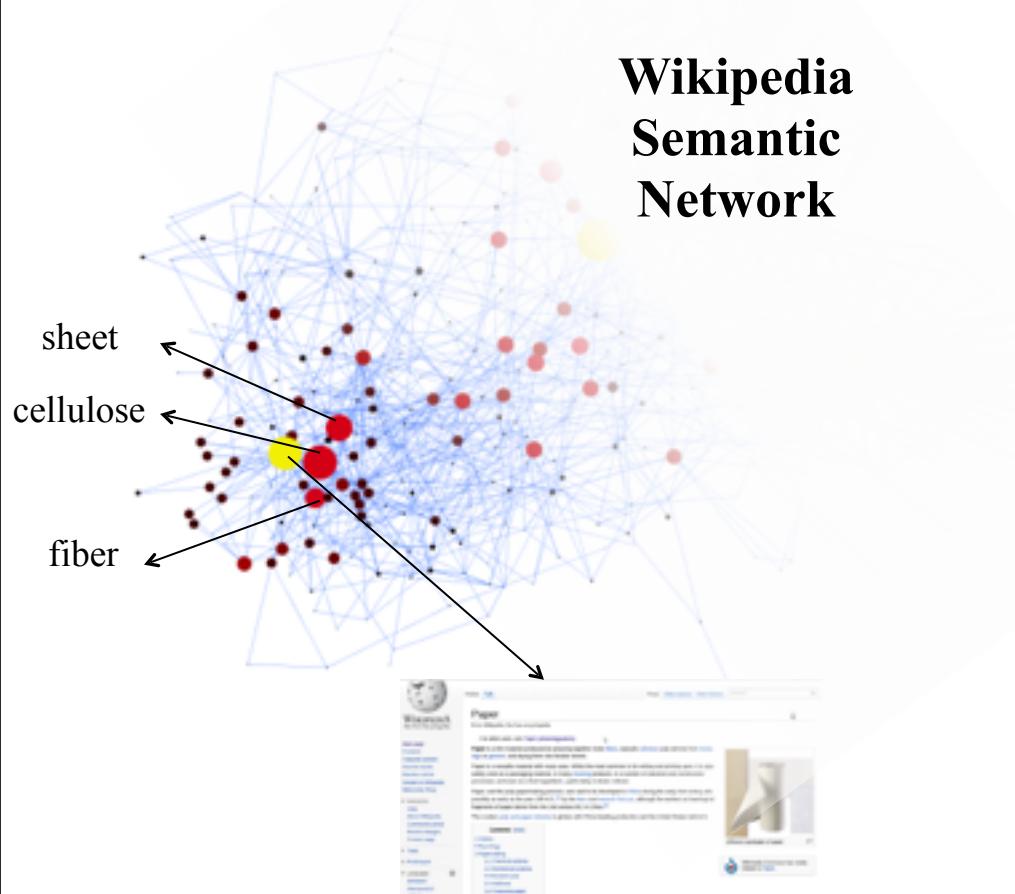
Definitional similarity (OW-WN)

The screenshot shows the OmegaWiki page for the term "paper". The page header includes links for "Expression" and "Discussion", and a top bar with "English", "Create account", and "Log in" buttons. A message for anonymous users is displayed: "As an anonymous user, you can only add new data. If you would like to also modify existing data, please create an account and indicate your languages on your user page." The main content starts with the word "paper" in bold. Below it, language options include "Other languages: Basque, Breton, Castilian, Catalan" and "Language: English". Under "Adjective", there is a single entry: "▶ paper : Made of paper." Under "Substantive", there are two entries highlighted with yellow boxes:

- ▶ paper : Felted or matted sheets of cellulose fibers, formed on a fine-wire screen from a dilute water suspension, and bonded together as the
- ▶ paper : A scholarly written work describing the results of observations or stating hypotheses.

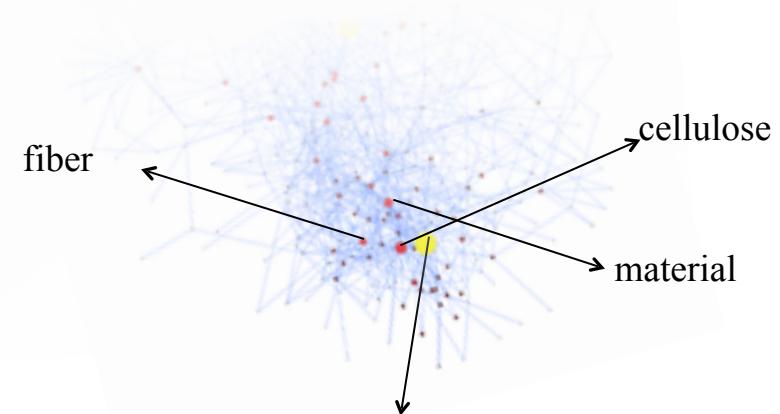
1. **paper** -- a material made of cellulose pulp derived mainly from wood or rags or certain grasses.
2. **paper** -- a n essay (especially one written as an assignment).
3. **paper** -- a daily or weekly publication on folded sheets; contains news and articles and advertisements.

Structural similarity



**Wikipedia
Semantic
Network**

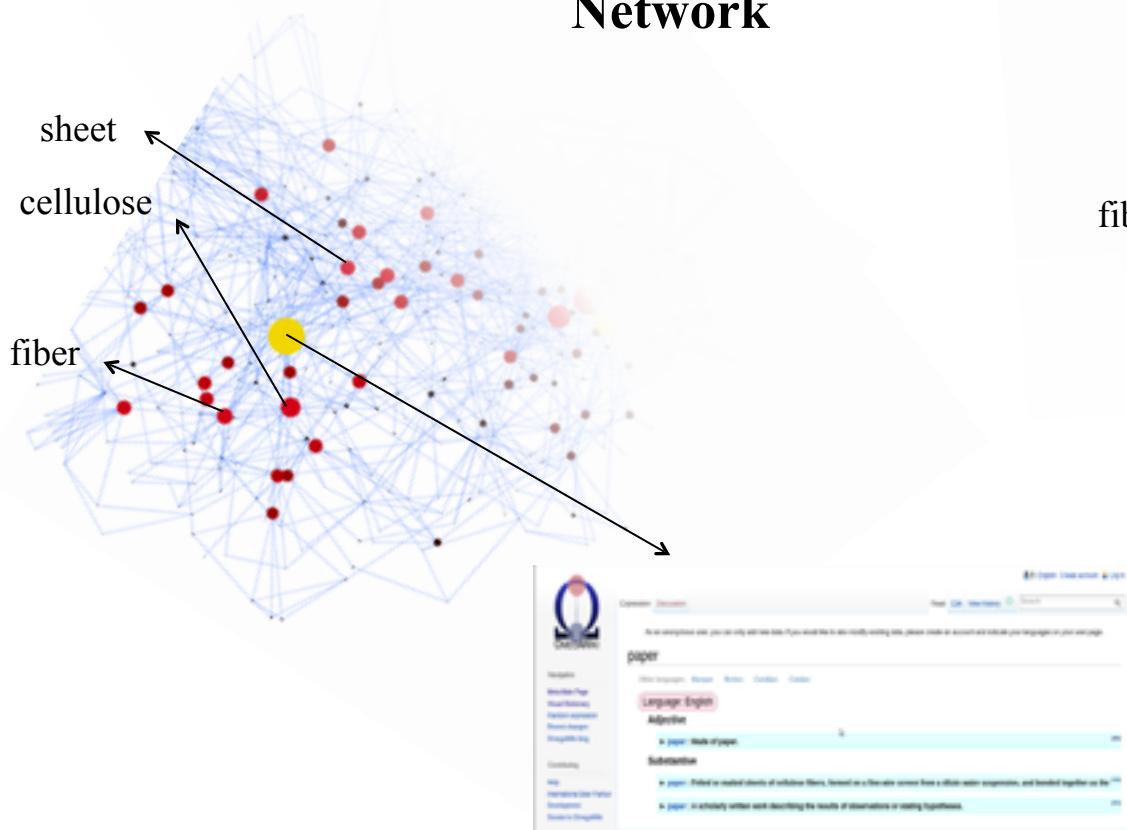
**WordNet
Semantic
Network**



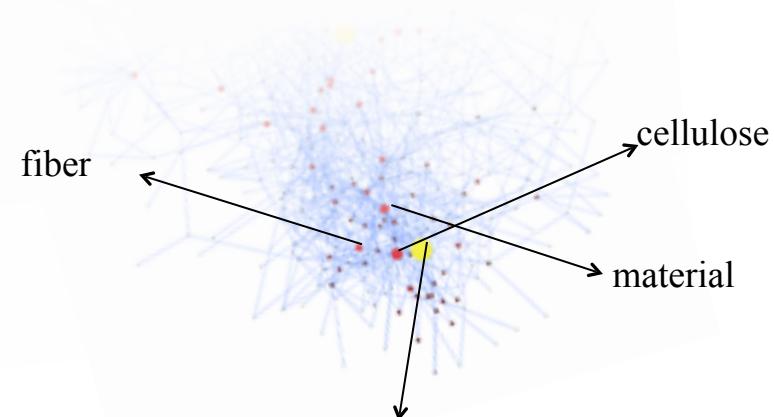
1. **paper** -- a material made of cellulose pulp derived mainly from wood or rags or certain grasses.

Structural similarity

OmegaWiki Semantic Network



WordNet Semantic Network



1. **paper** -- a material made of cellulose pulp derived mainly from wood or rags or certain grasses.

BabelNet 1.1.1 → 2.0

1. From **six** languages to **50** languages;
2. From **two** resources to **four**;
3. From **5** million synsets to **9** million concepts;
4. From **12** million lexicalizations to **15** million lexicalizations;
5. From **140** million semantic relations to **250** million semantic relations.

BabelNet 2.0 is online: <http://babelnet.org>

[about](#) • [statistics](#) • [preferences](#)



A very large multilingual encyclopedic dictionary and ontology

BabelNet is both a multilingual encyclopedic dictionary, with lexicographic and encyclopedic coverage of terms in 50 languages, and an ontology which connects concepts and named entities in a very large network of semantic relations, made up of more than 9 millions of entries. [Read more...](#)



The BabelNet API

```
BabelNet bn = BabelNet.getInstance();
System.out.println("SYNSETS WITH English word: \"bank\"");
List<BabelSynset> synsets = bn.getSynsets(Language.EN, "bank");
for (BabelSynset synset : synsets)
{
    System.out.print(" =>(" + synset.getId() + ") SOURCE: " + synset.getSource() +
                    "; WN SYNSET: " + synset.getWordNetOffsets() + ";\n" +
                    " MAIN LEMMA: " + synset.getMainLemma() + ";\n SENSES (German): { ");
    for (BabelSense sense : synset.getSenses(Language.DE))
        System.out.print(sense.toString() + " ");
    System.out.println("}\n -----");
    Map<IPointer, List<BabelSynset>> relatedSynsets = synset.getRelatedSynsets();
    for (IPointer relationType : relatedSynsets.keySet())
    {
        List<BabelSynset> relationSynsets = relatedSynsets.get(relationType);
        for (BabelSynset relationSynset : relationSynsets)
        {
            System.out.println("   EDGE " + relationType.getSymbol() +
                               " " + relationSynset.getId() +
                               " " + relationSynset.toString(Language.EN));
        }
    }
    System.out.println(" -----");
}
```

Retrieve all synsets with the English lemma “bank”

Print information about each synset

Get the (relation, synsets) map of the synset neighbours

Get the synsets related by a given relation type

Print the information of each related synset

BabelNet goes at a faster pace than I can cope with

Key fact!



Anatomy of BabelNet 2.0

Previous version had 6!

50 languages covered (including Latin!)

List at <http://babelnet.org/stats.jsp>



Anatomy of BabelNet 2.0

50 languages covered (including Latin!)

9.3M Babel synsets (concepts and named entities)

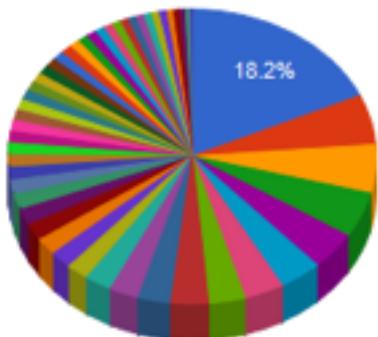
50M word senses

262M semantic relations (28 edges per synset on avg.)

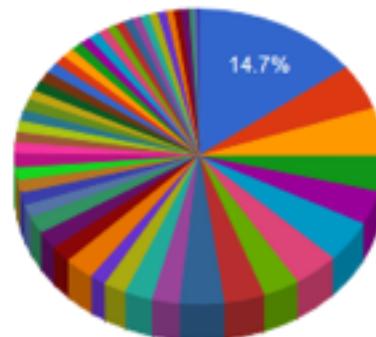
7.7M synset-associated images

18M textual definitions

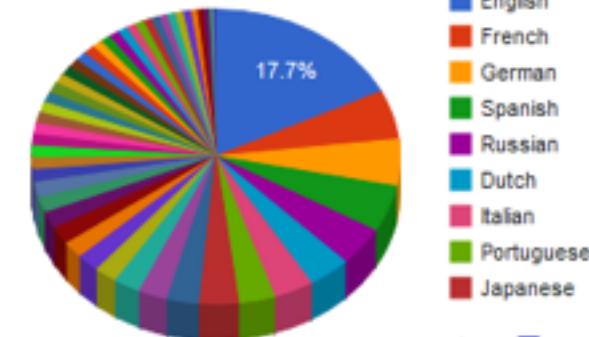
Lemmas by Language



Synsets by Language



Senses by Language



Anatomy of BabelNet 2.0

50 languages covered (including Latin!)

Integrates:

WordNet 3.0

Wikipedia (2012 dump)

OmegaWiki: a collaborative multilingual dictionary

Open Multilingual WordNet [Bond and Foster, 2013]

Translations for all open-class parts of speech

WordNet+OpenMultilingualWordNet+Wikipedia+...

Meaning: airplane¹ • ID: bn:00001697n • Type: Concept [\[explore\]](#)

Senses:

- W airplane¹, aeroplane¹, plane¹
- W aereo, aeroplano, apparecchio, areoplano, aeroplano, avión, aeroplà, aviò, samolot, aeroplan, エアプレイン, エアプレーン, 機, 航空機, 銀翼, 飛行機, 鵬翼, جنگل, اوپر, fly, flygar, flyver, lentokone, kapal terbang, pesawat, pesawat terbang, aeroplan, fly, flyvemaskine, flyver, kapal terbang, pesawat, pesawat terbang, avion, aeroplan, aeroplano, avión, hegazkin, aireplano, aireko, abioi
- W Fixed-wing aircraft, Aeroplano, Avión, Flugzeug, Avion, Aviò, Samolot, Avião, Vliegtuig, Самолёт, Avion, Uçak, 飛行機, 固定翼飛機, Máy bay, Самолет, طفارة, Flygplan, اوپر, مروحيات, वायुयान, Літак, Fly, Lentokone, Letoun, Repülőgép, 비행기, Pesawat bersayap tetap, Flyvemaskine, Авион, Lietadlo, Lėktuvas, Pesawat kepak kaku, Avion, Αεροπλάνο, Vliegtuig, Letalo, Aeroplani, Lennuk, Eittleán, Lidmašīna, Awyren, Flugvél, Ndege (uanahewa), Eroplano, Hegazkin, Aviadilo, Avión (medio de transporte), Aéroplanum
- W Air-plane, Fixed wing, Airplane, Airoplane, Æroplane, Aeroplane, Fixed-winged aircraft, Aeroplane, Fixed-wing, Airplanes, Aeroplanes, Plane (aircraft), Aereoplano, Aeroplani, Aereo, Aeroplano, Avión de alas fijas, Aviones, Aeronave de alas fijas, Aeronave de ala fija, Avion de alas fijas, Viajes aereos, Viajes aéreos, Aëroplan, Flugmaschine, Verkehrsluftzeug, Starrflügel, Avions, Fonctionnement d'un avion, Aeronau d'ala fixa, Aeroplà, Avlons, Aparat latający, Aeroplani, Samoloty, Cechy konstrukcyjne samolotów, Aviões, Aeronave de asa fixa, Aeroplano, Avião a Jato, Jatos, Jato, Avioneta, Indústria aeronáutica, Aviao, Vliegtuigen, Vliegmachine, Аэроплан, Самолёты, Самолет, Самолеты, Avioane, Teyyare, Sabit kanatlı uçak, Uçak kalkış-iniş, Tayyare, 大型飛行機, 大型旅客機, 蒸氣飛行機, 電氣動力式飛行機, 固定机翼飞机, 飞机, 固定翼航空器, 固定翼機, 固定翼, 固定翼机, 飛機, 固定翼飞机, 定翼機, Tàu bay, Máy bay cánh cố định, Phi cơ, Самолети, Аероплан, Aircraft, الطوارق, الطوارق, الطوارق, الطوارق, الطوارق, الطوارق

+OmegaWiki+automatic translations...

Q aeroplane, plane, fixed-wing aircraft, bird, aereo, aeroplano, avión, aeroplano, Flugzeug, avion, aéroplane, samolot, aeroplano, avião, toestel, vliegtuig, vliegmachine, kist, аэроплан, самолёт, avion, uçak, 飛行機, 飛機, 飞龙机, 飞行机, 飞机, 飛龍機, máy bay, аероплан, самолет, طائرة, flygplan, օլոր, هواپیما, वायु-यान, හැවා ජාහාස, аероплан, літак, fly, lentokone, letadlo, repülőgép, 비행기, fly, flyvemaskine, авион, lietadlo, lėktuvas, aeroplanas, kapal terbang, avion, zrakoplov, αεροπλάνο, vliegtuig, letalo, avion, aeroplan, lennuk, aeroplaan, eitleán, awyren, flugvél, eroplano, hegazkin, aviadilo, flugmašlino, avio, aeroplanum

Y velivoli ad ala fissa, ala fissa, aereo, aerei, aeronaves de ala fija, ala fija, avión, aviones, starrflügler, flugzeug, flugzeuge, flugzeugen, aéronefs à voilure fixe, voiture fixe, avion, avions, aeronaus d'ala fixa, ala fixa, avió, avions, stalej skrzydło, samolotów, aeronaves de asa fixa, de asa fixa, asa fixa, avião, aviões, vliegtuigen met vaste vleugels, vliegtuig, vaste vleugels, vliegtuigen, самолетов, самолеты, avion, cu aripă fixă, avioane, sabit kanatlı uçak, sabit kanatlı, uçak, 固定翼, 飛行機, 固定翼飞机, 固定翼, 飞机, máy bay cánh cố định, cánh cố định, bay, máy bay, самолет, с неподвижно крило, неподвижно крило, самолети, الطائرات, flygplan med fasta vingar, fasta vingar, flygplan, պլ գլուխ, مهاری, ക്രാഫ്റ്റ് കുമ്പാം, සං කුඩා කුඩා, fast ving fly, faste vinger, fly, fast ving, lentokone, fixovaný-letadla křídla, pevné křídlo, letadel, letadla, merevszárnyú repülőgépek, merevszárnyú, repülőgép, repülőgépek, 고정 날개, pesawat sayap tetap, sayap tetap, pesawat, pesawat terbang, fastvingede fly, fastvingede, fly, фиксне крила авиона, фиксне крила, авиона, fixovaný-letadla krídla, lietadie, lietadlo, fiksuotujų sparnų orlaivis, fiksuoto sparno, fiksuotujų sparnų, lėktuvo, lėktuval, pesawat sayap tetap, sayap tetap, kapal terbang, nepokretnim krilima zrakoplova, zrakoplova, αεροσκάφη σταθερής πτέρυγας, αεροπλάνο, σταθερής πτέρυγας, αεροπλάνο, vaste-vlerk vliegtuie, vastevlerk, vliegtuig, vaste-vlerk, vliegtuie, nepremičnimi krili letala, fiksno krilo, letala, aeroplan, aeroplanē, fikseeritud hävitusalennuk, jäigatiivaliste, lennuk, lennukid, eitleán, sciatháin dhobhogha,

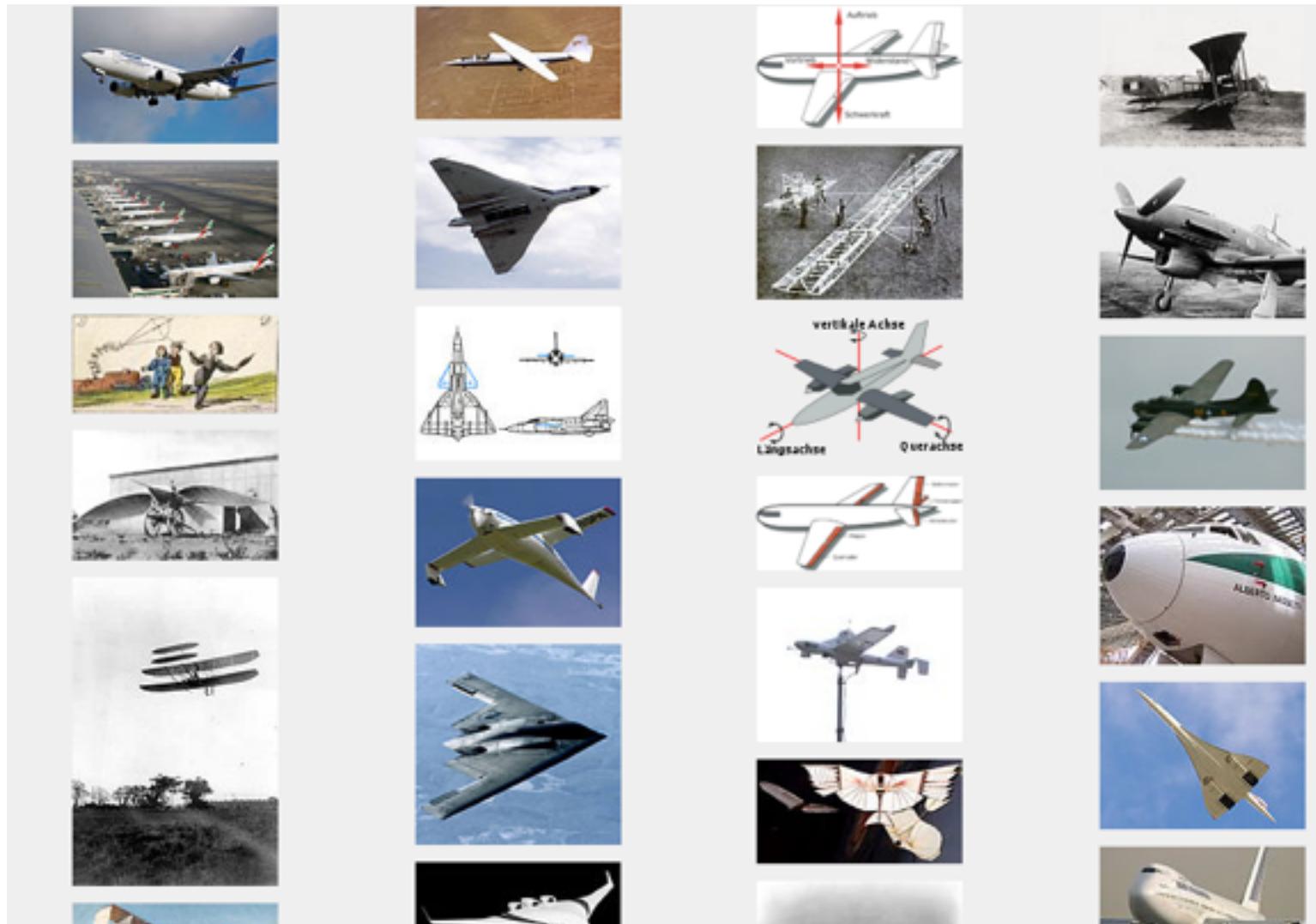
+textual definitions

Glosses:	
Q A powered heavier-than-air aircraft with fixed wings that obtains lift by the Bernoulli effect and is used for transportation.	
W an aircraft that has a fixed wing and is powered by propellers or jets; "the flight was delayed due to trouble with the airplane"	
W A fixed wing aircraft is an aircraft capable of flight using wings that generate lift due to the vehicle's forward airspeed and the shape of the wings.	
Q Veicolo dotato di ali che grazie all'effetto Bernoulli riesce, sollevandosi da terra, a rimanere in aria e viene utilizzato per il trasporto.	
W L'aeroplano, o aereo, è un mezzo di trasporto.	
Q Artefacto más pesado que el aire con alas fijas que puede levantarse del suelo debido al efecto de Bernoulli y es utilizado para transporte.	
W Avión, también denominado aeroplano, es un aerodino de ala fija, o aeronave con mayor densidad que el aire, provisto de alas y un espacio de carga capaz de volar, impulsado por uno o más motores.	
W Ein Flugzeug ist ein Luftfahrzeug, das schwerer als Luft ist und den zum Flug nötigen dynamischen Auftrieb mit starren Tragflächen erzeugt.	
Q Un engin aérien motorisé plus lourd que l'air pourvu d'ailes fixes lui assurant une portance en vertu de l'effet de Bernoulli, qui est utilisé à des fins de transport.	
W Un avion, selon la définition officielle de l'Organisation de l'aviation civile internationale, est un aéronef plus lourd que l'air, entraîné par un organe moteur, dont la sustentation en vol est obtenue principalement par des réactions aérodynamiques sur des surfaces qui restent fixes dans des conditions données de vol.	
W Un avió, és una aeronau, més pesant que l'aire, capaç de volar per l'atmosfera.	
W Um avião ou aeroplano é qualquer aeronave que necessita de asas fixas para se sustentar no ar.	
Q Een aangedreven vliegtuig die zwaarder is dan lucht met vaste vleugels die lift krijgt door het Bernoulli-effect en gebruikt wordt voor transport.	

+Wikipedia categories

Categories: Aeronautics, Aircraft configurations, Aeroplani, Aeronáutica, Configuraciones de aeronaves, Inventos, Luftfahrzeug, Type d'aéronef, Avions, Samoloty, Vliegtuig, Uçaklar, Fransız icatları, 航空, 航空器, 航空器种类, 飛機, 大氣污染, Máy bay, Hàng không, Khí cự bay, Vận tải, Phương tiện giao thông, Profile khí động lực học, Lực nâng máy bay, Самолети, Обществен транспорт, الملاحة, تشكيلات الطائرات, طائرات, كلمات من وضع مجمع اللغة العربية, اختراعات أمريكا, هوايماء, فناوری هوایی, فناوری هوانی, هواپیمای رادارگریز, هوانوردی, विमानन, वायु परिवहन के साधन, वायुयान, यान, Літаки, Повітряні судна, Fly, Kollektivtrafikk, Lentokoneet, 1900-luvun keksinnöt, Letadla, Repülögépek, Repüléstechnika, 항공기, Aeronautika, Pesawat, Konfigurasi Pesawat, Flyvemaskiner, Luftfart, Maskiner, Авиони, Воздухоплови, Lietadlai, Orlaivai, Zrakoplovi, Vliegtuile, Zrakoplovi, Letala, Teknologji, Aeronautikē, Komunikacion, Lennundus, Lennuliiklus, Aerárthach, Awyrənnau, Flugvélar, Chomboanga, Aeronautika, Aviadiloj, Transportaj rimedoj, Aviado, Aviación, Aerinavigatio, Res volantes

+images



Evaluations: I have to go fast here!



WordNet-Wikipedia mapping accuracy

Overall quality of the mapping: ~84%

On a random sample of 1k Wikipages

Note: this concerns only those 50k synsets in the intersection

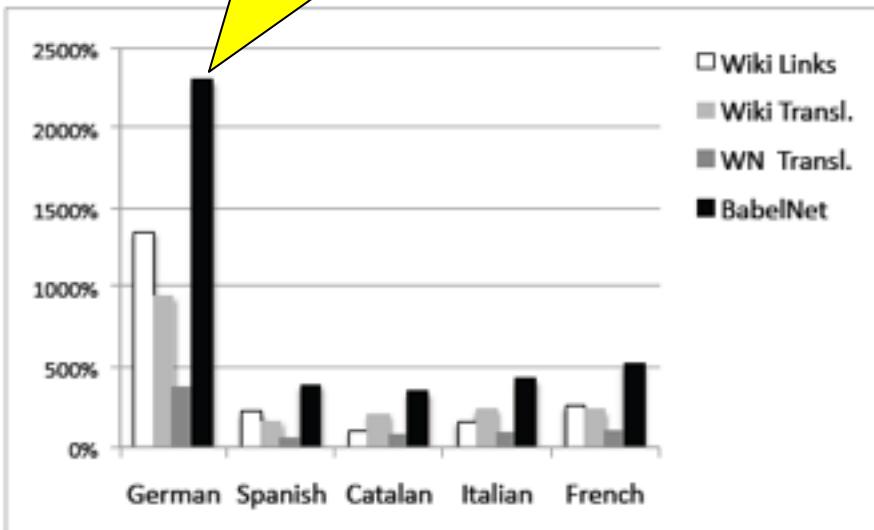
Quality of the mapping of frequent words: ~91%



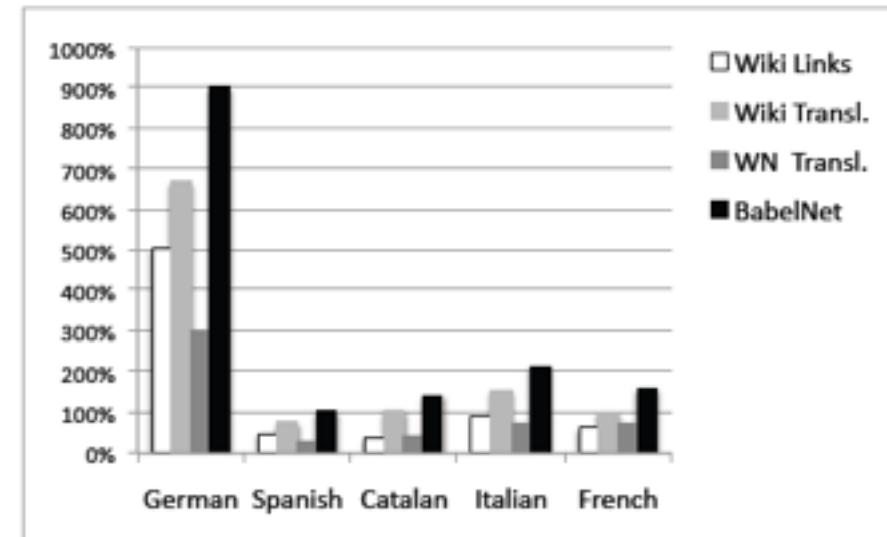
Evaluation of BabelNet against gold standard resources

Up to +2300% new senses!

Extra-coverage



(a) word senses



(b) synsets

ONE DOES NOT SIMPLY USE BABELNET



imgflip.com

Navigli

BabelNet goes to the Multilingual
Semantic Web: ESWC 2014 Tutorial

Part 3: Identifying multilingual concepts and entities in text

82³

BabelNet as a Multilingual Inventory for

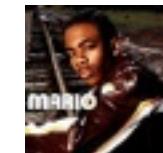
- Word Sense Disambiguation



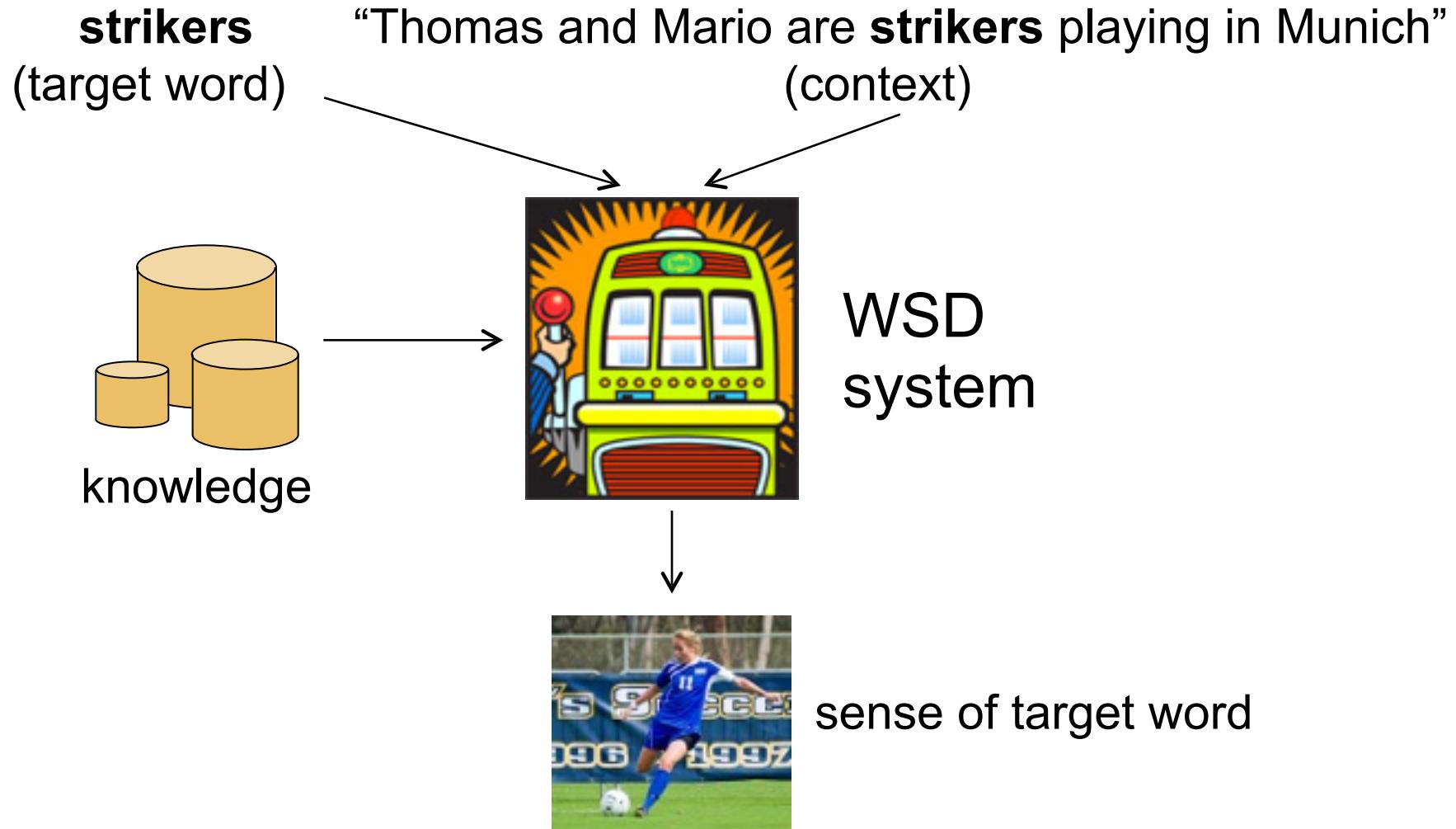
The concept of soccer can be lexicalized in Italian as calcio, in Spanish as fútbol, and in Dutch as Voetbal

- Entity Linking

The text *mario* can be used to represent different things such as the video game character or a soccer player or even a music album



Word Sense Disambiguation in a Nutshell



Main references

A complete survey of the field (course reference):

Navigli R. Word Sense Disambiguation: a Survey. ACM Computing Surveys, 41(2), ACM Press, 2009, pp. 1-69.

WSD book:

Agirre E. and Edmonds P. Word Sense Disambiguation: Algorithms and Applications, New York, USA, Springer, 2006.

Another survey from last decade:

Ide N. and Véronis J. Word Sense Disambiguation: The State of The Art. Computational Linguistics, 24(1), 1998, pp. 1-40.

WSD: main approaches

Supervised WSD

Frames the problem as a classification task

Relies on hand-labeled training sets

Knowledge-based WSD

Uses knowledge resources to identify the best senses for words in context

Typically, it does not need a training phase and relies on an existing inventory of senses

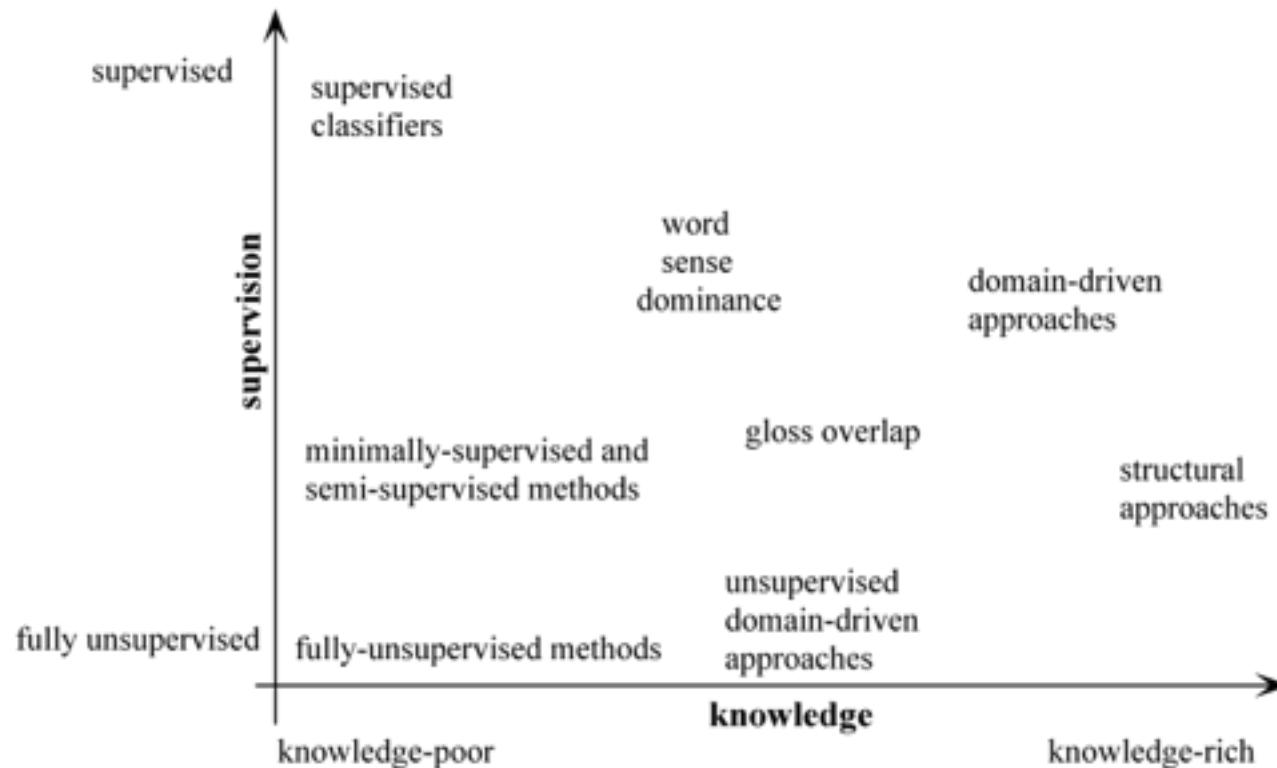
Word Sense Discrimination / Induction

Unsupervised WSD: clustering

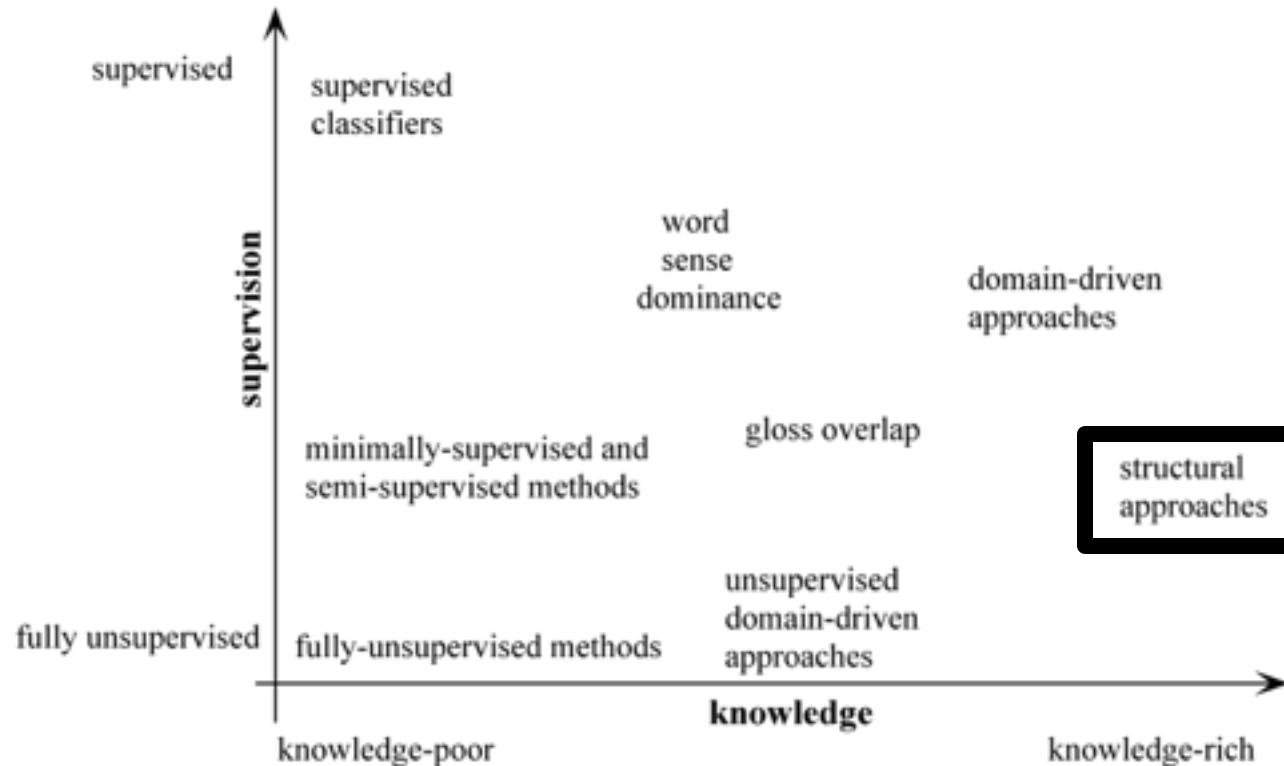
Does not need manually-tagged datasets

Can make the task more difficult to evaluate

Supervision: labeled data vs. knowledge



Supervision: labeled data vs. knowledge



Knowledge-based WSD: structural approaches

Structural approaches analyze and exploit the structure of a knowledge resource.

Given a knowledge resource (e.g. BabelNet):

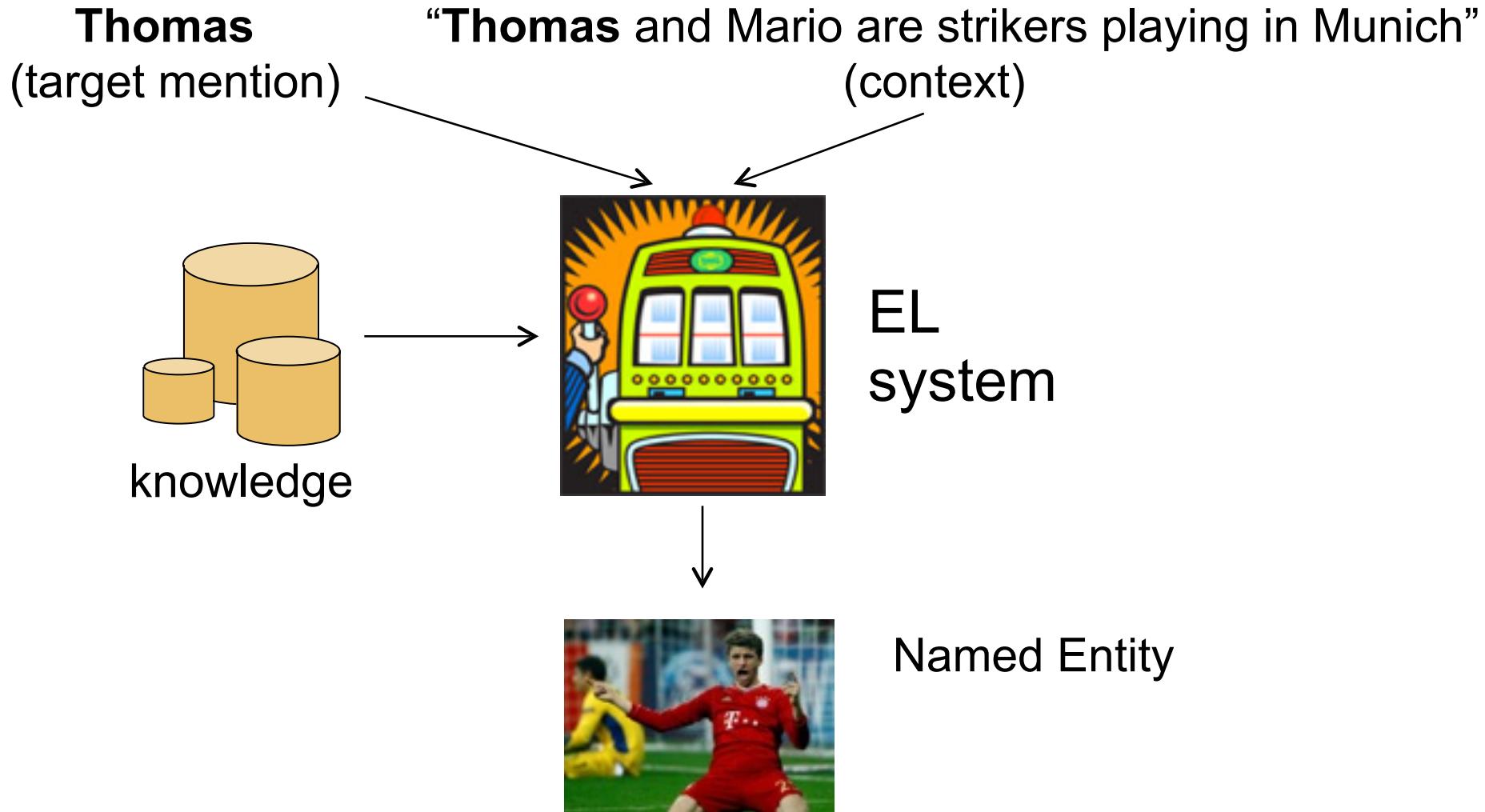
View the resource as a *graph*

Apply a method that *makes use of the structure of the graph*

State-of-the-art WSD systems

- Supervised:
 - It Makes Sense: a **SVM** trained on manually annotated corpora;
- Structural:
 - UKB: an application of the **Personalized PageRank** on semantic networks containing word senses

Entity Linking in a Nutshell



Entity Linking

EL encompasses a set of similar tasks:

- **Named Entity Disambiguation**, that is the task of linking entity mentions in a text to a knowledge base
- **Wikification**, that is the automatic annotation of text by linking its relevant fragments of text to the appropriate Wikipedia articles.

Entity Linking

State-of-the-art approaches are based on the following concepts:

- **Collective disambiguation** of mentions vs. independent disambiguation of mentions;
- **Enforcing semantic coherence** among the chosen named entities;
- **Efficiency**: there are different orders of magnitude between the number of word senses and named entities!

State-of-the-art EL systems

- **AIDA**: a graph-based framework for the exploitation of similarity measures between candidate entities;
- **KORE**: a graph-based similarity measure integrated with the key phrases contained within the context to disambiguate entities;
- **Tagme**: a combination of the Milne-Witten measure (hyperlinks similarity on Wikipedia) with the commonness of an entity;
- **Wikifier**: a global and local approach based on the TF-IDF score combined with hyperlinks in Wikipedia;
- **DBpedia Spotlight**: a generative model based on counts about the manually disambiguated Wikipedia hyperlinks.

State-of-the-art EL systems

- **AIDA**: a graph-based framework for the exploitation of **similarity measures** between candidate entities;
- **KORE**: a graph-based **similarity measure** integrated with the key phrases contained within the context to disambiguate entities;
- **Tagme**: a combination of the Milne-Witten measure (**hyperlinks similarity on Wikipedia**) with the commonness of an entity;
- **Wikifier**: a global and local approach based on the TF-IDF score combined with **hyperlinks in Wikipedia**;
- **DBpedia Spotlight**: a generative model based on counts about the manually disambiguated **Wikipedia hyperlinks**.

A Joint approach to WSD and EL

- Knowledge-based approaches perform well on both these two tasks:
 - The main difference is the kind of inventory used



A Joint approach to WSD and EL



BabelNet is a multilingual inventory for both word senses and named entities!

Babelfy: A Joint approach to WSD and EL [Moro et al., 2014]

Personalized PageRank is the state-of-the-art method for graph-based word sense disambiguation, however it **cannot** be run for each new input on **huge graphs**.

Idea: Precompute semantic signatures for the nodes!

Semantic signatures are the most **relevant nodes** for a given node in the graph computed by using **random walk with restart**

Andrea Moro and Alessandro Raganato and Roberto Navigli. 2014. Entity Linking meets Word Sense Disambiguation: a Unified Approach. TACL

<http://babelfy.org>

Semantic Signatures: RWR

1. Start from one target vertex of the semantic netowork;
2. Randomly select a neighbor of the current vertex or restart at the target vertex;
3. Keep counting the hitting frequencies;
4. Take the most visited vertices.

Semantic Signatures

offside#n#1

striker#n#1



athlete#n#1



soccer_player#n#1

sport#n#1



A Joint approach to WSD and EL

1. Given an input text select all the possible candidate meanings from BabelNet by matching mentions with BabelNet lexicalizations;
2. Connect all the candidate meanings by using semantic signatures;
3. Extract a dense subgraph containing semantically coherent candidates;
4. Select the most connected candidate for each fragment of text.

Step 1: Find all possible meanings

Thomas and Mario are strikers playing in Munich

Seth Thomas



Mario (Character)



striker (Sport)



Munich (City)



Thomas Müller



Mario (Album)



Striker (Video Game)



FC Bayern Munich



Mario Gómez



Striker (Movie)



Munich (Song)



Step 1: Find all possible meanings

Thomas and Mario are strikers playing in Munich

Seth Thomas



Mario (Character)



striker (Sport)



Munich (City)



Thomas Müller



Mario (Album)



Striker (Video Game)



FC Bayern Munich



Mario Gómez

Thomas (novel)



Striker (Movie)

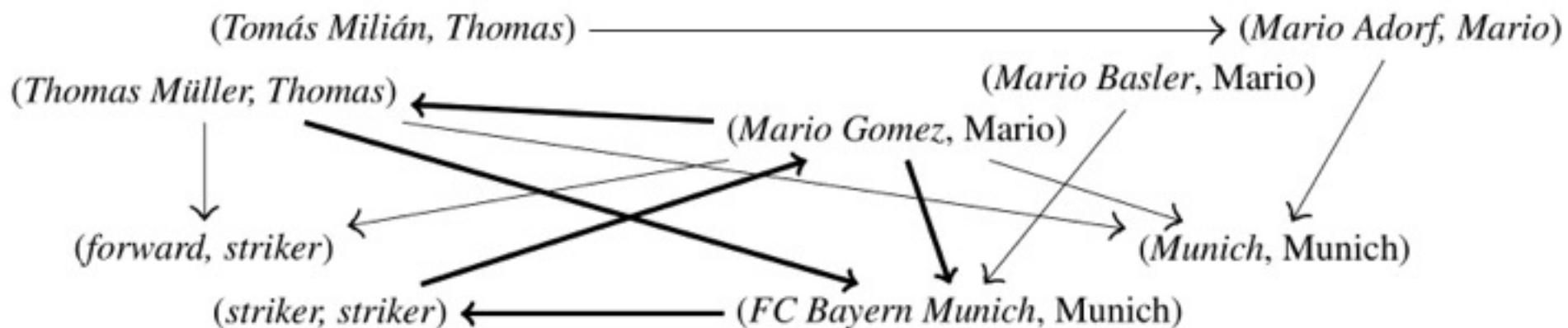


Munich (Song)



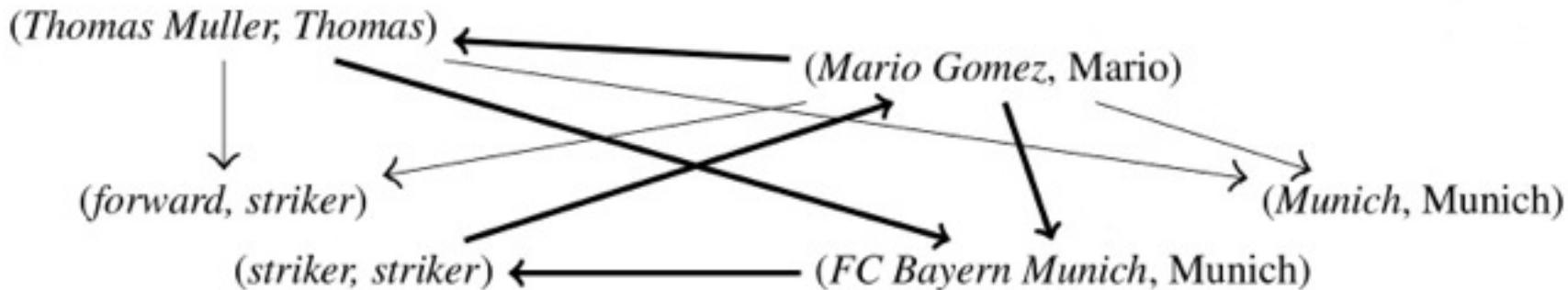
Step 2: Connect all the candidate meanings

Thomas and Mario are **strikers** playing in Munich



Step 3: Extract a dense subgraph

Thomas and Mario are **strikers** playing in **Munich**



Step 4: Select the most reliable meaning

Thomas and Mario are strikers playing in Munich

Seth Thomas



Mario (Character)



striker (Sport)



Munich (City)



Thomas Müller



Mario (Album)



Striker (Video Game)



FC Bayern Munich



Mario Gómez



Striker (Movie)



Munich (Song)



Thomas (novel)



Experimental Results

System	Sem07	Sens3	SemEval-2013 English			French		German		Italian		Spanish	
	WN	WN	WN	Wiki	BN								
Babelfy	62.7	67.7	65.9	87.4	69.2	71.3	*56.9	*79.0	69.4	83.3	66.6	82.8	69.5
IMS	62.1	69.2	65.7	—	—	—	—	—	—	—	—	—	—
UKB w2w	*56.0	*65.3	61.3	—	60.8	—	60.8	—	66.2	—	67.3	—	70.0
UMCC-DLSI	—	—	64.7	54.8	68.5	*60.5	60.5	58.1	62.8	*58.3	65.8	*61.0	71.0
DAEBAK!	—	—	—	—	60.4	—	53.8	—	59.1	—	*61.3	—	60.0
GETALP-BN	—	—	51.4	—	58.3	—	48.3	—	52.3	—	52.8	—	57.8
MFS	65.8	69.6	*63.0	*80.3	*66.5	69.4	45.3	83.1	*67.4	82.3	57.5	82.4	*64.4

System	F1
(Ponzetto and Navigli, 2010)	85.5
Babelfy	84.6
UoR-SSI	84.1
UKB w2w	83.6
NUS-PT	*82.3
MFS	77.4

Experimental Results

System	KORE50	CoNLL
Babelfy	71.5	82.1
KORE-LSH-G	64.6	81.8
KORE	63.9	*80.7
MW	*57.6	82.3
Tagme	56.3	70.1
KPCS	55.6	82.2
KORE-LSH-F	53.2	81.2
UKB w2w (on BabelNet)	52.1	71.5
Illinois Wikifier	41.7	72.4
DBpedia Spotlight	35.4	34.0

Where to go from here?

We are now able to attach both meanings and entities to a natural language expression.

What can we do with that?

Why not extracting semantic relations between them!

Part 4:

Adding multilingual semantics to Information Extraction

Motivation (part 1): Open Information Extraction

(Nunavut, is a territory of, Canada)

1. Automatically read arbitrary text;
2. extract meaningful information;
 1. (subject, relational phrase, object)
3. a unified knowledge base.



is located in



is a member of



is a territory of



Motivation (part 1): Open Information Extraction

Pros:

- Open set of relations
- Web-scale extraction

No semantics

Cons:

- Textual information
- Shallow syntax

Part-of-speech tags

Motivation (part 2): Knowledge Acquisition

A **machine readable** semantically well-defined representation of knowledge:

- A set of semantically well-defined concepts;
- A set of semantically well-defined relations;



Administrative Area: The smallest administrative area in which this area is contained.



is an **administrative area** of

territory

territory



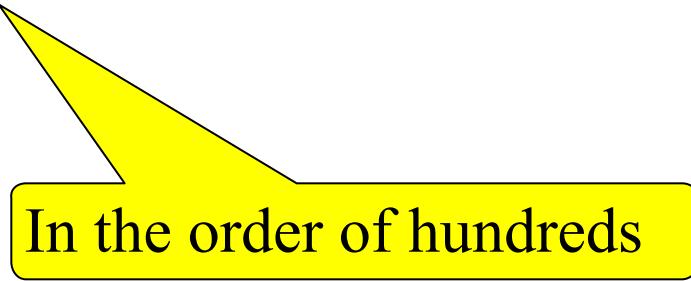
Motivation (part 2): Knowledge Acquisition

Pros:

- Semantically well-defined Relation
- Semantically well-defined Concepts
- High coverage for the considered relation

Cons:

- Partial representation of knowledge
 - Limited number of semantic relation



In the order of hundreds

Our work in context

Systems	Semantic Relations	Open set of relations	Syntax
OIE (e.g. Reverb)			Shallow
KA (e.g. YAGO2)			Shallow
WiSeNet http://lcl.uniroma1.it/wisenet			Shallow/ Dependency

A. Fader, S. Soderland and O. Etzioni. Identifying Relations for Open Information Extraction. In Proc. of EMNLP, 2011.

J. Hoffart et al. YAGO2: A spatially and temporally enhanced knowledge base from Wikipedia. Journal of Artificial Intelligence, 2012.

A. Moro and R. Navigli, WiSeNet: Building a Wikipedia-based Semantic Network with Ontologized Relations. In Proc. of CIKM , 2012.

A. Moro and R. Navigli, Integrating Syntactic and Semantic Analysis into the Open Information Extraction Paradigm. In Proc. of IJCAI, 2013.

Input of our Information Extraction System



Nunavut is a territory of **Canada**.

Nunavut is located in the **Canadian Arctic**.



***Taproot Theatre Company is a member of
Theatre Communications Group***



***Wiesbaden State Library is funded by the
State of Hesse and located in Wiesbaden.***

Step 1: Relation Extraction

Argument

Relational phrase

Argument

(Nunavut, is a territory of, Canada)

(Nunavut, is located in the, Canadian Arctic)

Relation
Instance

**(Taproot Theatre Company,
is a member of,
Theatre Communications Group)**

Filtering out bad relational phrases

Wiesbaden State Library is funded by the **State of Hesse** and located in **Wiesbaden**.

Filtering out bad relational phrases

Wiesbaden State Library is funded by the **State of Hesse** and located in **Wiesbaden**.

(**Wiesbaden State Library**, is funded by the, **State of Hesse**)

Filtering out bad relational phrases

Wiesbaden State Library is funded by the **State of Hesse** and located in **Wiesbaden**.

(**Wiesbaden State Library**, is funded by the, **State of Hesse**)



Filtering out bad relational phrases

Wiesbaden State Library is funded by the **State of Hesse** and located in **Wiesbaden**.

(**Wiesbaden State Library**, is funded by the, **State of Hesse**)



(**Wiesbaden State Library**, is funded by the State of Hesse and located in, **Wiesbaden**)

Filtering out bad relational phrases

Wiesbaden State Library is funded by the **State of Hesse** and located in **Wiesbaden**.

(**Wiesbaden State Library**, is funded by the, **State of Hesse**)

(**Wiesbaden State Library**, is funded by the State of Hesse and located in, **Wiesbaden**)



Filtering out bad relational phrases

Wiesbaden State Library is funded by the **State of Hesse** and located in **Wiesbaden**.

(**Wiesbaden State Library**, is funded by the, **State of Hesse**)

(**Wiesbaden State Library**, is funded by the State of Hesse and located in, **Wiesbaden**)

(**State of Hesse**, and located in, **Wiesbaden**)

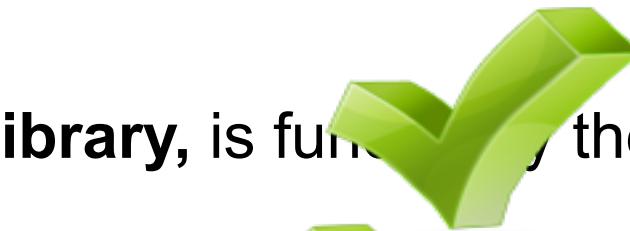
Filtering out bad relational phrases

Wiesbaden State Library is funded by the **State of Hesse** and located in **Wiesbaden**.

(**Wiesbaden State Library**, is funded by the, **State of Hesse**)

(**Wiesbaden State Library**, is funded by the State of Hesse and located in, **Wiesbaden**)

(**State of Hesse**, and located in, **Wiesbaden**)

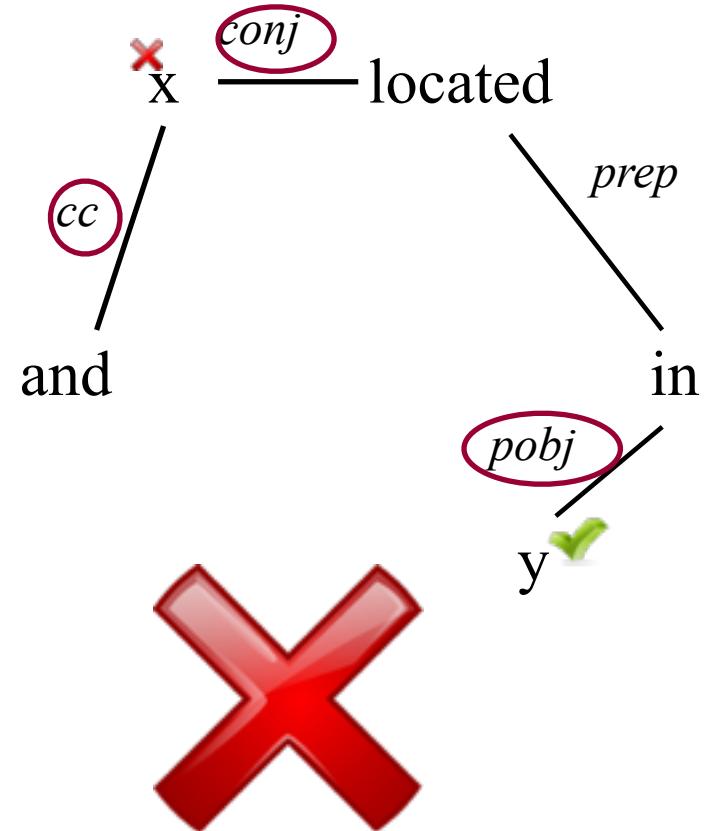
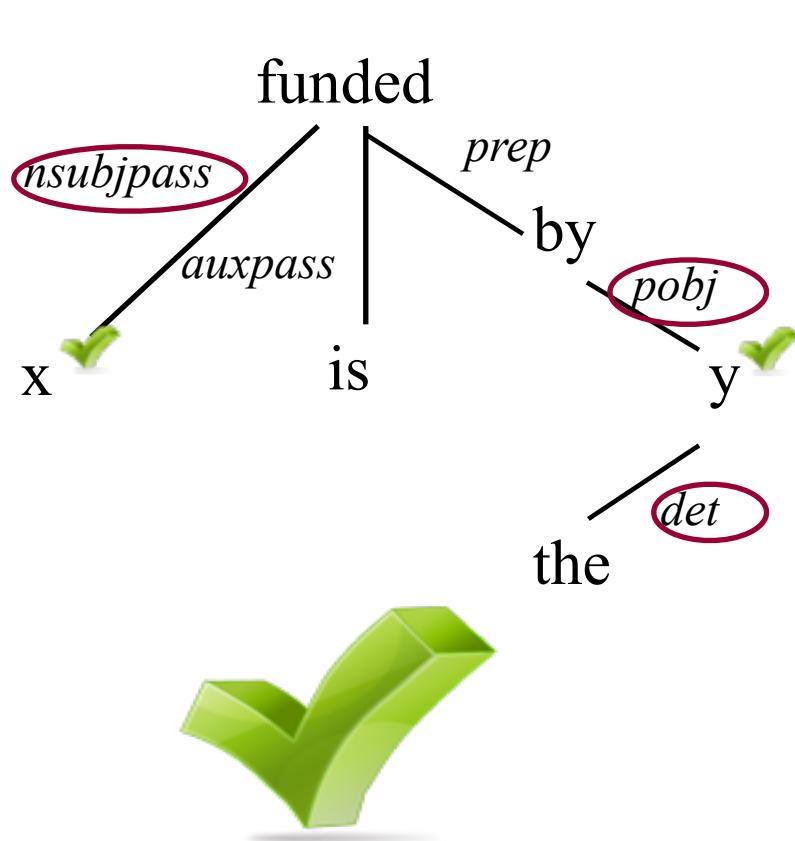


Filtering out bad relational phrases

Artificial relational phrase

x is funded by the y

x and located in y



Filtering out bad relations

- We do not parse the whole corpus
- We parse short text
- We keep only syntactically sound relational phrases
- We can now exploit the syntactic structure of the relational phrases

Highly-scalable and easier for a multilingual setting !!!

Syntactic-grounded

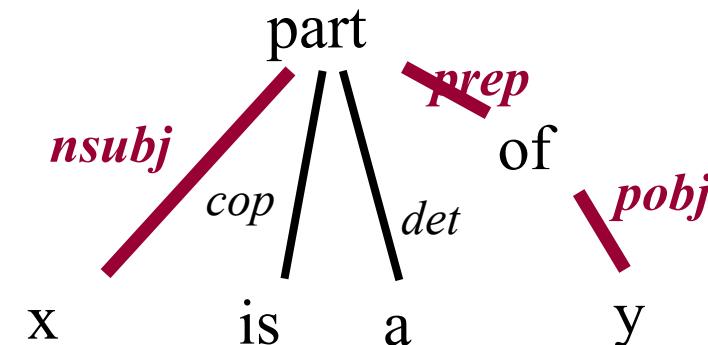
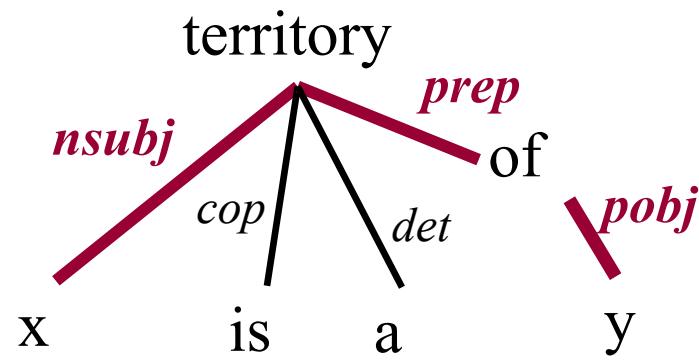
Step 2: Relation Ontologization

We take into account **two** aspects:

1. The **syntactic structure** of the relational phrases;
2. The **distributional similarity** of the relational phrases' words

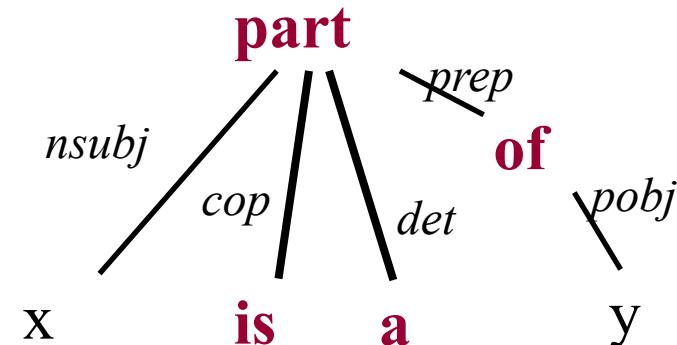
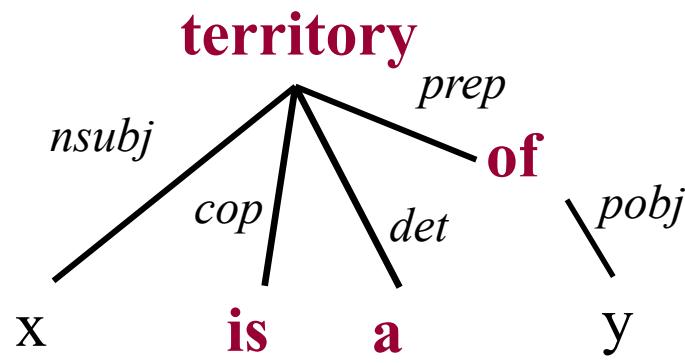
Step 2: Relation Ontologization

Syntactic Assumption: Two **relational phrases** can represent the **same semantic relation type** only if they share the **same syntactic shortest path** between X and Y



Step 2: Relation Ontologization

Distributional Assumption: Two **relational phrases** can represent the **same** semantic relation type only if they use **semantically similar words**.



Soft Clustering to build Relation Synsets

{is a part of, is a territory of, ..., is a province of}

{is a part of, is a member of, ..., is an element of}

{made her acting debut in the, made his professional debut in the, ..., made his acting debut in the}

{used to build used to construct , ..., used to manufacture}

Soft Clustering to build Relation Synsets

{**is a part of**, is a territory of, ..., is a province of}

{**is a part of**, is a member of, ..., is an element of}

{made her acting debut in the, made his professional debut in the, ..., made his acting debut in the}

{used to build used to construct , ..., used to manufacture}

Semantic Type Signatures

Our **semantic type signatures** are distributional vectors.

1. The top two levels of the Wikipedia Categories Hierarchy

Academia, ..., Biology, ..., Geography, ..., Law, ..., Politics

2. The extracted **relation instances** to compute **frequencies**

(Nunavut, is a territory of, Canada)



Geography



Geography

Semantic Type Signatures

Domain	Relation Synset	Range
Arts	{is located in the small village of, . . ., is located in the small rural town of}	Places
Corporate groups	{is a member of an, . . ., were the members of the}	Corporate groups
Geography	{is a valley of, is a zone of, . . ., is a territory of}	Geography

Relation Type Disambiguation

Ambiguous Relation Instance:		
<i>(Natural language processing,</i>	<i>is a field of,</i>	<i>Computer science)</i>

Relation Type Disambiguation

Ambiguous Relation Instance:		
<i>(Natural language processing,</i>	<i>is a field of,</i>	<i>Computer science)</i>
Relation Synset Candidates:		
{Subfields by academic discipline, ..., Science}	{is a field of, is an area of, is studied in}	{Scientific Disciplines, ..., Science}
{Agriculture, ..., Horticulture and gardening}	{is a field of, is cultivated with, where grows}	{Fruit, ..., Cultivars}
{Cities, ..., Villages}	{is a field of, was the site of, is the battlefield, }	{Battles, ..., Wars}

Relation Type Disambiguation

Ambiguous Relation Instance:		
<i>(Natural language processing,</i>	<i>is a field of,</i>	<i>Computer science)</i>

Relation Synset Candidates:		
{Subfields by academic discipline, ..., Science}	{is a field of, is an area of, is studied in}	{Scientific Disciplines, ..., Science}
{Agriculture, ..., Horticulture and gardening}	{is a field of, is cultivated with, where grows}	{Fruit, ..., Cultivars}
{Cities, ..., Villages}	{is a field of, was the site of, is the battlefield, }	{Battles, ..., Wars}

Semantic Relation:		
(Natural language processing,	{is a field of, is an area of, is studied in},	Computer science)

Part 5: Multilingual Lexical Knowledge as Linked Data

Context



- MultiJEDI project (<http://multijedi.org>)

Multilingual Joint word sense Disambiguation

- objective 1: create a large-scale multilingual knowledge base

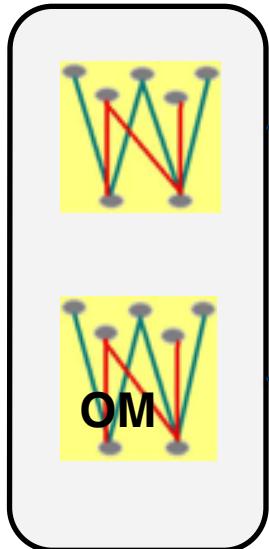


- objective 2: enable multilingual text understanding

<http://babelfy.org>

Acquisition

Traditional lexical resources



Collaborative lexical resources



- fully-structured
- manually curated by experts
- available for a few languages
- difficult to maintain and update

- semi-structured
- collaboratively built by the crowd
- highly multilingual
- up-to-date

BabelNet

- a lexical-semantic resource
 - a multilingual dictionary, with encyclopedic and lexicographic entries
 - a semantic network, with concepts and named entities connected through semantic relations
- acquisition
 - automatic integration of several resources
- structure/main components
 - Babel synsets: represent a given meaning
 - Babel senses: terms in different languages
 - semantic relations

Structure

Babel synset

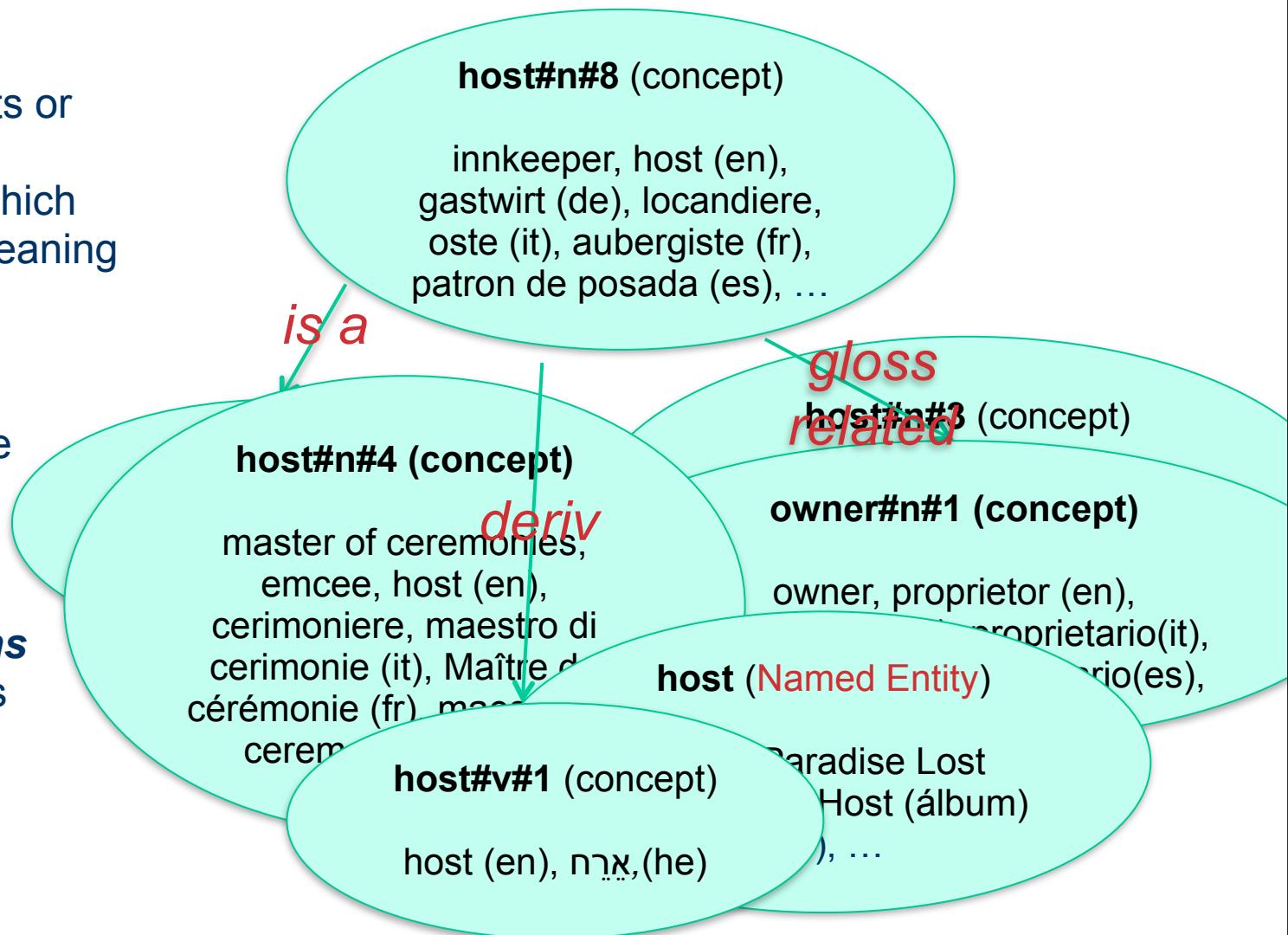
- encode concepts or named entities
- group senses which express their meaning

Babel senses

- terms in multiple languages

Semantic relations

- connect synsets



Information

- **Concepts** host#n#4
- **Senses/Translations**

master of ceremonies, emcee, host (en), ceremoniere, maestro di ceremonie (it),
Maître de cérémonie (fr), maestro de ceremonias (es), ...

- **Definitions**

a person who acts as host at formal occasion (en)

le terme maître de cérémonie désigne un animateur de spectacle... (fr)

Der MC oder älter Master of Ceremonies ist im englischen Sprachraum ... (de)

- **Categories**

Entertainment, Television presenters, Radio presenters, ...

- **Relations**

lexical and semantic relations from WordNet and Wikipedia

- concept
- senses
- definitions
- categories
- semantic relations
- metadata

Meaning: master of ceremonies¹ • ID: bn:00030515n • Type: Concept

Senses:

- master of ceremonies", emcee", host"
- cerimoniere, maestro di ceremonie, presentatore, ホスト, 司会, 司会者, 進行係, 進行係り, 進行員, 進行掛り, maestro de ceremonias
- Master of Ceremonies, MC (音楽), Maître de cérémonie, MC (Musik), Master of Ceremonies, MC (Hip Hop), Master of Ceremonies, MC (ヒップホップ), Master of Ceremonies, Maestro de ceremonias
- Wack rapper, Master of Ceremony, Master of ceremonies, Emcee, Wack mc, Wack MCs, Masters or Mistresses of Ceremonies, Sucker MC, Master Of Ceremony, Emcee, Wack MC, MC, MCing, Wack emcee, Sucker mc, Mistress of Ceremonies, M.C., Mistress of ceremonies, Host (radio), Chair (MC), Wack mcs, MVC, Compering, M. C., Host (event), Wack rappers, M. C., MC (音楽), Master of Ceremony, Emcee, Maître de Cérémonie, Maître De Cérémonie, Maître de ceremonie, Maître de cérémonie, Maître de Cérémonie, Master of Ceremonies, MCing, MC (Hip Hop), Emcee, MCing, マイクロフォン・コントローラ,マイクロフォン・コントローラー, Эм-ок, Эм-ок, Мастер вечеринок, Эмси, MC, Maestra de ceremonias
- host, presentador
- المنظم, 主持人, hôte, σκηνοθέτης, ٹوبوچی, gestgjafi, host, ホスト, anfitrón
- emcee, المتربي, 司仪, 主持人, maître des cérémonies, emcee, zemeremonienmeister, телестарڈт, emcee, ผู้ดำเนินรายการ, skipstjóri vígslu, emcee, maestro delle ceremonie, emcee, 司会者, церемониймейстер, ведущий, maestro de ceremonias, emcee

Glosses:

- A moderator or master of ceremonies for a performance.
- a person who acts as host at formal occasions (makes an introductory speech and introduces other speakers)
- A Master of Ceremonies, or Compère, is the official host of a staged event or similar performance.
- Au sens strict, le terme maître de cérémonie désigne un animateur de spectacle, c'est-à-dire la personne qui dirige une ou des cérémonies, une fête, une soirée ou un spectacle.
- Der MC oder älter Master of Ceremonies ist im englischen Sprachraum generell der Moderator einer Bühnenshow oder einer anderen Performance.
- Master of Ceremonies, noto anche con l'acronimo MC, è un'espressione di lingua inglese che significa letteralmente "maestro di ceremonie".
- MC, или 'эм-ок' в рэп-культуре и хип-хопе — артист, в сопровождении электронной танцевальной музыки произносящий со сцены слова — заранее сочинённые или импровизированные, обычно в виде рэпа — чтобы раззадорить публику, а также представить диджея.
- Moderador o maestro de ceremonias de una presentación.
- Maestro de ceremonias es el encargado de presentar una ceremonia o evento especial o un espectáculo.

Categories:

- Entertainment, Television presenters, Radio presenters, Broadcasting occupations, مهنة التلفزيون, مهن في التلفزيون, كليات من وظائف القطاع العربي
- Lexique du hip-hop, Animateur de télévision, Métier de la télévision, Rap, Hardstyle, Fernsehmoderator, Beruf (Kleinunterhalt), โทรทัศน์ผู้ดำเนินรายการ, Terminologia hip hop, Professioni della televisione, Conduttori televisivi, Ocupaciones, Presentadores de televisión, Locutores de radio.

DBpedia: Master of Ceremonies

babelnet.org#lemonRepresentation

- the RDF resource consists of a set of **Lexicons**, one per language.

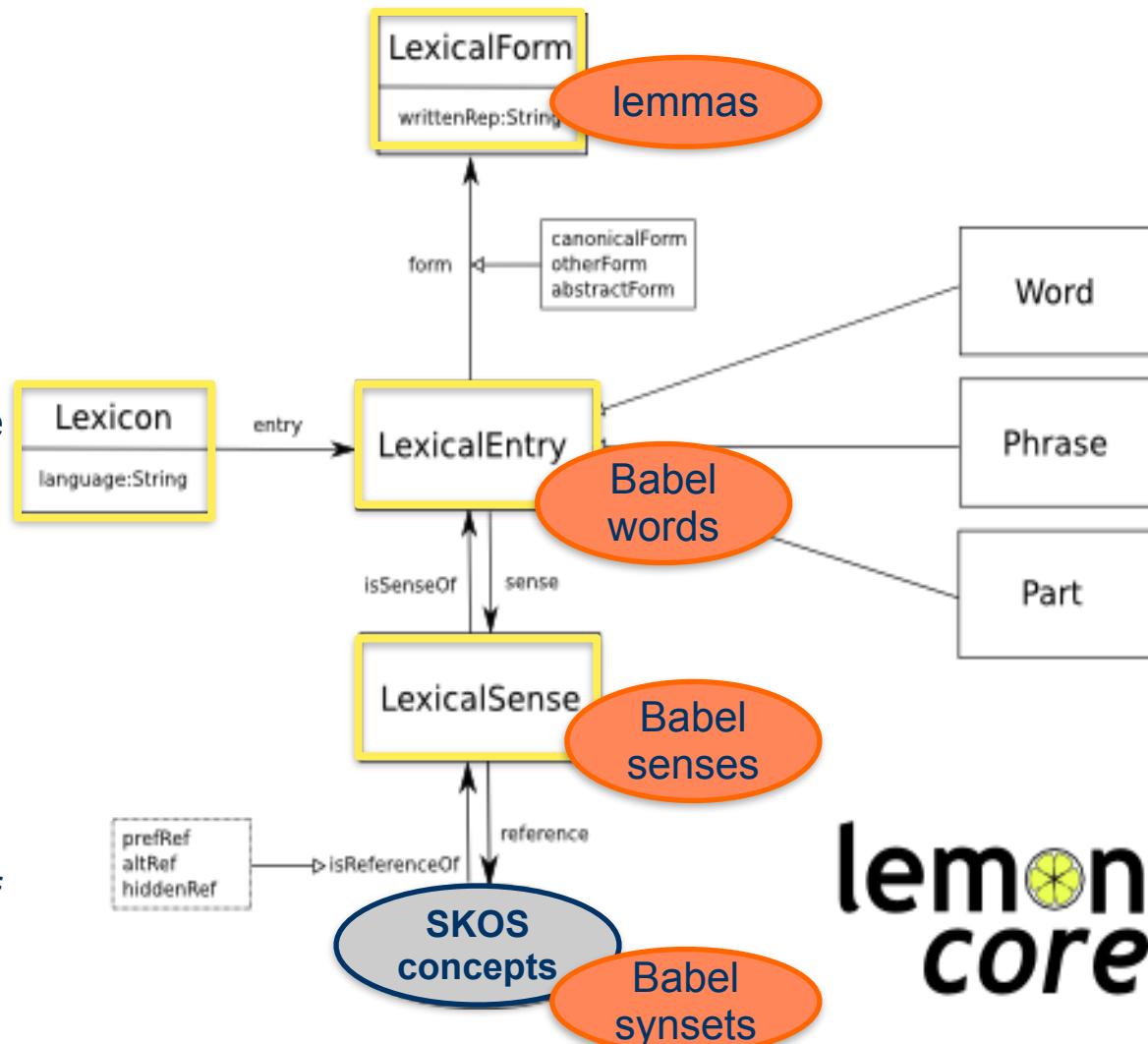
- Lexicons gather **Lexical Entries**

which comprise the forms of an entry;
in our case: *words* of the Babel lexicon.

- **Lexical Forms** encode the surface realisation(s) of Lexical Entries;
in our case: *lemmas* of Babel words.

- **Lexical Senses** represent the usage of a word as *reference* to a specific concept;
in our case: Babel senses.

- **Skos Concepts** represent ‘units of thought’;
in our case: Babel synsets.



lemon

- backbone of BabelNet lexical knowledge representation
- <http://www.lemon-model.net/lemon#>

SKOS (Simple Knowledge Organization System)

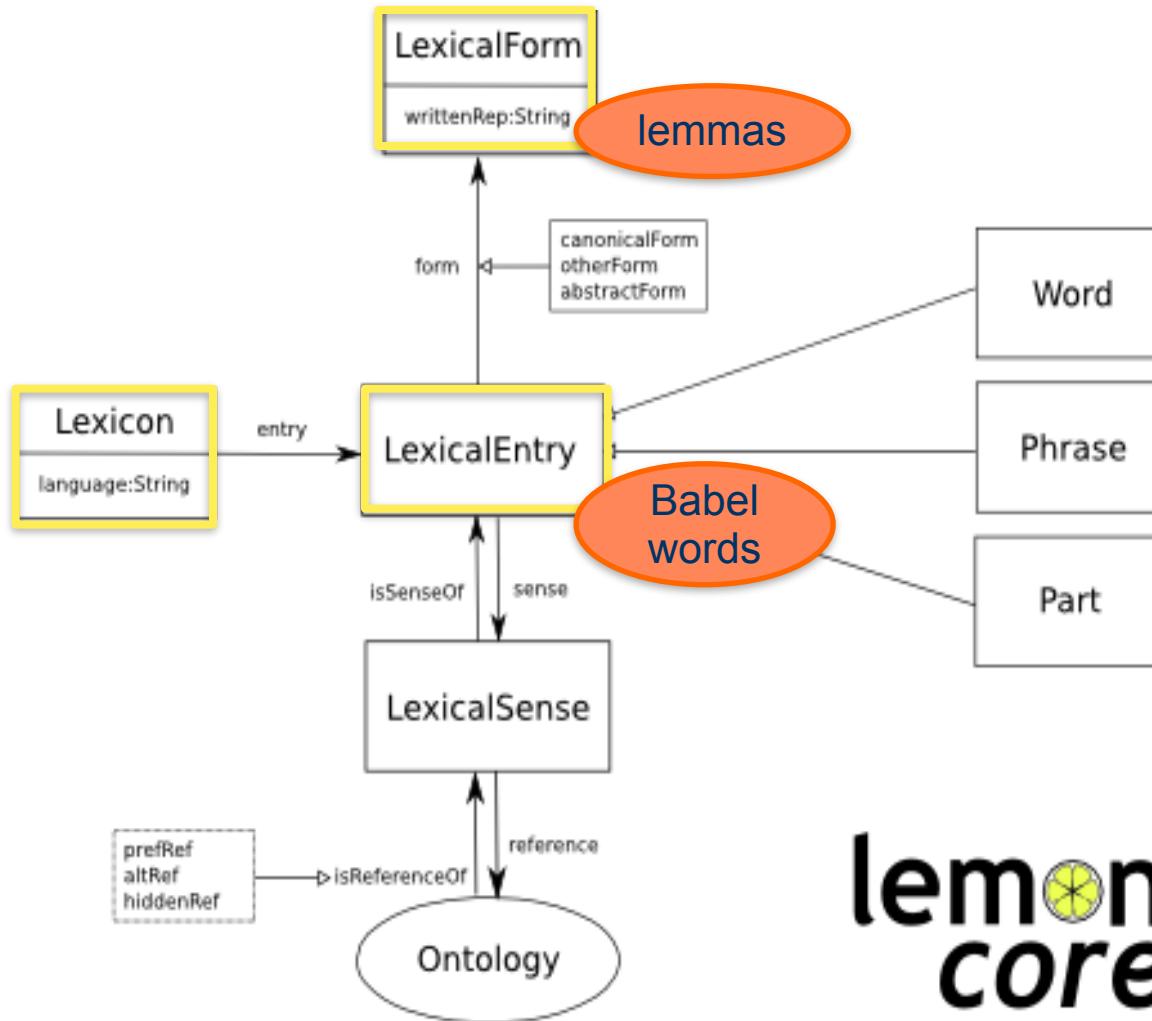
- skos:Concept class represents Babel synsets
- <http://www.w3.org/2004/02/skos/core#>

LexInfo 2.0

ontology which describes linguistic information
used here to represent various linguistic information
<http://www.lexinfo.net/ontology/2.0/lexinfo#>

BabelNet:

domain name: <http://babelnet.org/2.0/>
vocabulary: <http://babelnet.org/model/babelnet#>



lemon
core

babelnet.org#representingWords

language - EN

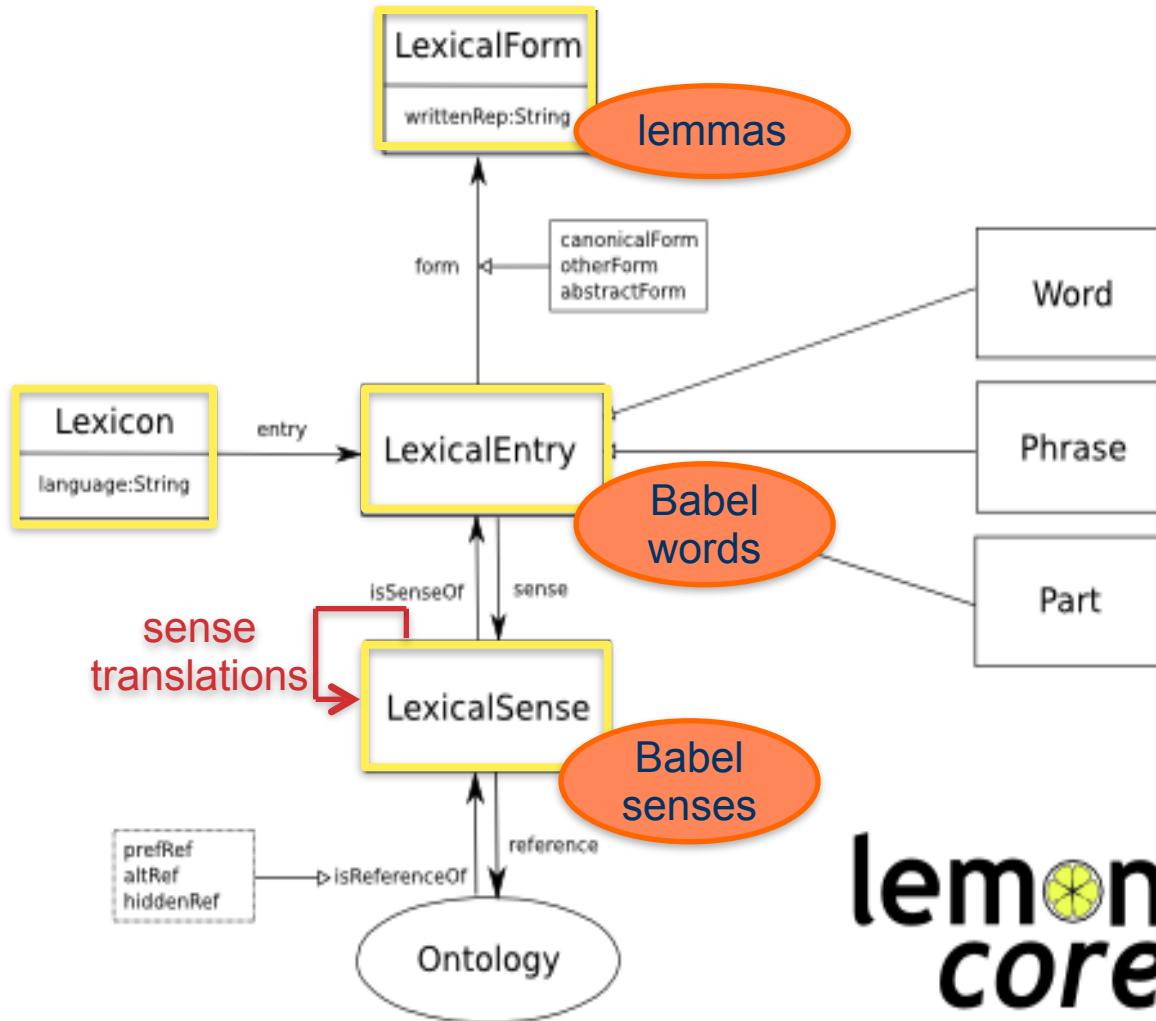
language - EN

```
@prefix lemon: <http://www.lemon-model.net/lemon#> .
@prefix bn: <http://babelnet.org/2.0/> .
@prefix lexinfo:<http://www.lexinfo.net/ontology/2.0/lexinfo#> .

bn:lexicon_EN
a          lemon:Lexicon ;
lemon:entry           bn:host_n_EN .

bn:host_n_EN
a          lemon:LexicalEntry ;
rdfs:label           "host"@en ;
lemon:canonicalForm bn:host_n_EN/canonicalForm ;
lemon:language        "EN" ;
lexinfo:partOfSpeech lexinfo:noun .

bn:host_n_EN/canonicalForm
a          lemon:Form ;
lemon:writtenRep    "host"@en .
```



lemon
core

babelnet.org#representingSenses

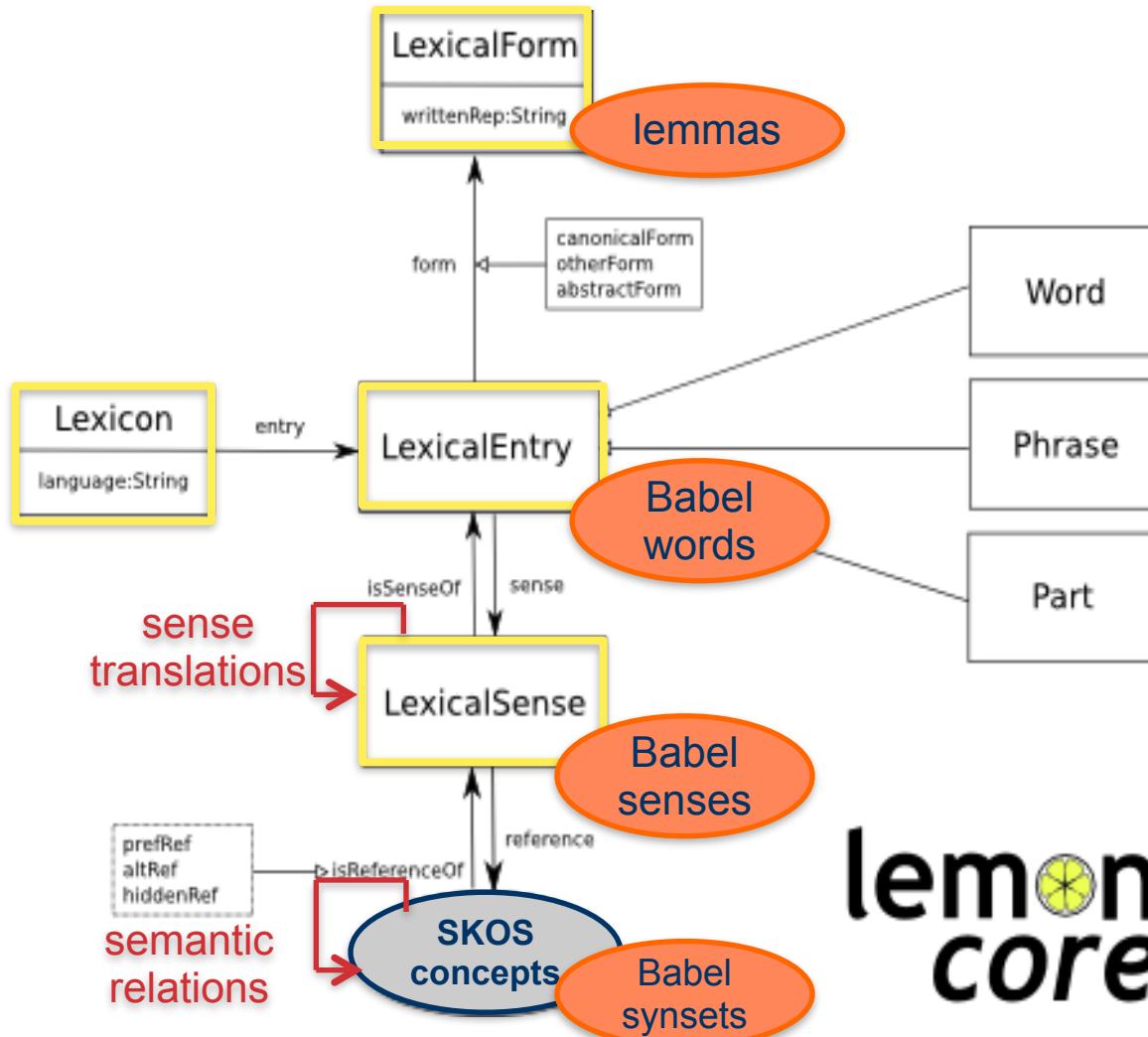
```
@prefix lemon: <http://www.lemon-model.net/lemon#> .  
@prefix bn: <http://babelnet.org/2.0/> .  
@prefix lexinfo: <http://www.lexinfo.net/ontology/2.0/lexinfo#> .
```

bn:host_n_EN

```
a                      lemon:LexicalEntry ;  
rdfs:label              "host"@en ;  
lemon:canonicalForm     bn:host_n_EN/canonicalForm ;  
lemon:language          "EN" ;  
lexinfo:partOfSpeech    lexinfo:noun ;  
lemon:sense              bn:host_EN/s00012029n .
```

http://babelnet.org/2.0/host_EN/s00012029n

```
a                      lemon:LexicalSense ;  
dc:source                http://wordnet.princeton.edu/  
dc:license                http://wordnet.princeton.edu/wordnet/license/  
lexinfo:translation      XXXXXX-TBC  
lemon:reference          bn:s00012029n
```



**lemon
core**

babelnet.org#representingSynsets

lemon:LexicalSense

lemon:reference =

```
http://babelnet.org/2.0/host_EN/s00012029n
a                      lemon:LexicalSense ;
dc:source      http://wordnet.princeton.edu/
dc:license     http://wordnet.princeton.edu/wordnet/license/
lexinfo:translation XXXXXX-TBC
lemon:reference bn:s00012029n

http://babelnet.org/2.0/s00030515n
a                      skos:Concept
bn-definition   http://babelnet.org/2.0/s00030515n_Gloss1_EN
dc:license      <http://creativecommons.org/licenses/by-nc-sa/3.0>

http://babelnet.org/2.0/s00030515n
a                      bn-lemon:BabelGloss
bn-lemon:gloss    'the owner or manager of an inn'
lemon:language    EN
dc:source        http://wordnet.princeton.edu/
dc:license       http://wordnet.princeton.edu/wordnet/license
```

babelnet.org#semanticRelations

http://babelnet.org/2.0/host_EN/s00012029n

```
a                      lemon:LexicalSense ;
dc:source      http://wordnet.princeton.edu/
dc:license      http://wordnet.princeton.edu/wordnet/license/
lexinfo:translation  XXXXXX-TBC
lemon:reference    bn:s00012029n
```

<http://babelnet.org/2.0/s00012029n>

```
a                      skos:Concept
bn-definition      http://babelnet.org/2.0/s00030515n_Gloss1_EN
dc:license        <http://creativecommons.org/licenses/by-nc-sa/3.0>
skos:broader       bn:s00061046n
lexinfo:derivedForm  bn:s00089484v
skos:related       bn:s00007078n
```

<http://babelnet.org/2.0/s00012029n>

```
a                      bn-lemon:BabelGloss
bn-lemon:gloss      'the owner or manager of an inn'
lemon:language      EN
dc:source      http://wordnet.princeton.edu/
dc:license      http://wordnet.princeton.edu/wordnet/license
```

Links towards encyclopedic datasets

Wikipedia categories and Dbpedia pages:

skos:exactMatch **on** skos:concept

Wikipedia pages:

bn-lemon:wikipediaPage **on** lemon:LexicalSense

DBpedia categories:

bn-lemon:dbpediaCategory **on** skos:concept

Links towards lexical datasets:

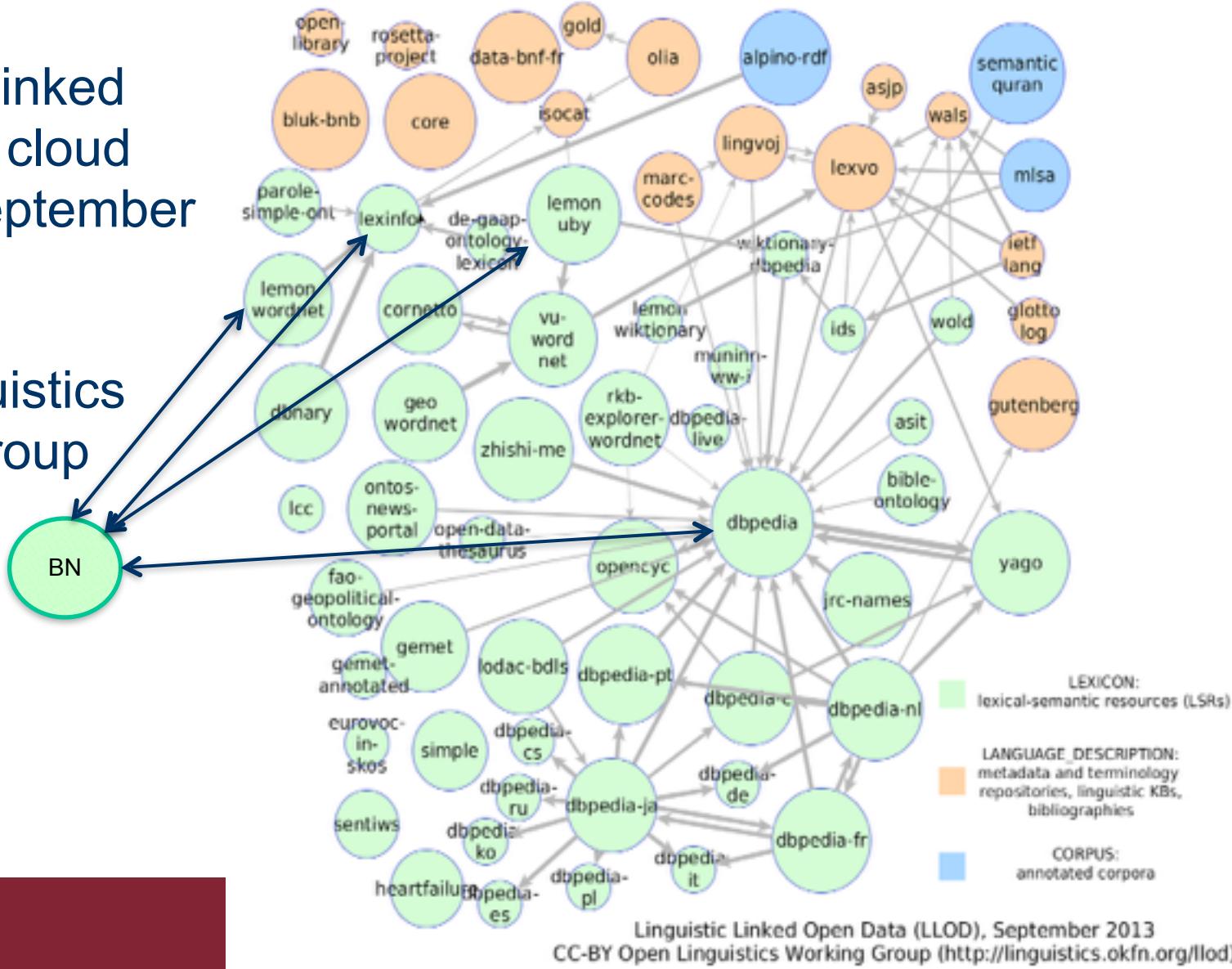
lemon-Wordnet: skos:exactMatch **on** skos:concept

WordNet RDF 3.1: skos:exactMatch **on** skos:concept

lemon-Uby, OmegaWiki (en): skos:exactMatch **on** skos:concept

babelnet.org#interlinking

- Linguistic Linked Open Data cloud (LLOD) September 2013
 - Open Linguistics Working Group



babelnet.org#publicationOnTheWeb

RDF dumps

- format: n-triples

- URIs and IRIs

- separated files per licenses

SPARQL endpoint

- virtuoso universal server

- <http://babelnet.org:8084/sparql/>

Dereferencing

- babelnet.org/2.0/

- Pubby Linked Data Frontend

- (<http://wifo5-03.informatik.uni-mannheim.de/pubby/>)

Possible additions:

- further morpho-syntactic information on lexical entries
- further links at sense level: Wiktionary
- more detailed information about translations

- multilingual ontology lexicalization
- multilingual ontology mapping
- linked data word sense disambiguation
- cross-resource validation of sense alignments
- ...

references/links

LIDER project: <http://www.lider-project.eu/>

lemon

J. McCrae et al. *Interchanging lexical resources on the Semantic Web*. Language Resources and Evaluation 46.4 (2012): 701-719.

website: <http://lemon-model.net/>

BabelNet

M. Ehrmann, F. Cecconi, D. Vannella, J. McCrae, P. Cimiano, R. Navigli.

Representing Multilingual Data as Linked Data: the Case of BabelNet 2.0, LREC 2014, Reykjavik.

R. Navigli and S. Ponzetto. *BabelNet: The Automatic Construction, Evaluation and Application of a Wide-Coverage Multilingual Semantic Network*, Artificial Intelligence, 193 (2012): 217-250.

website: <http://babelnet.org> , linked data: <http://babelnet.org/2.0/>

references/links

SKOS

A. Miles, Alistair and J. R. Pérez-Agüera. *SKOS: simple knowledge organisation for the web*. Cataloging & Classification Quarterly 43.3-4 (2007): 69-83.

<http://www.w3.org/2004/02/skos/>

lexInfo

P. Buitelaar, P. Cimiano, P. Haase & M. Sintek, M.. *Towards linguistically grounded ontologies*. In *The semantic web: research and applications* (2009): 111-125. Springer Berlin Heidelberg.

website: <http://lexinfo.net/>

Linked Data 4 rules

<http://www.w3.org/DesignIssues/LinkedData.html>

references/links

Linked Open Vocabularies (LOV)

<http://lov.okfn.org/dataset/lov/>

Virtuoso small how-to:

<https://confluence.deri.ie:8443/display/webstar/Virtuoso+hints+and+tips>

<https://confluence.deri.ie:8443/display/webstar/The+complete+tutorial+for+RDF+data+ingestion+in+Virtuoso>

http://babelnet.org/2.0/merci_n_FR !

Property Value

lemon:canonicalForm bn: merci_n_FR/
rdfs:label canonicalForm
lemon:language ‘merci’ @FR
lexinfo:partOfSpeech FR
lemon:sense lexinfo: noun
bn: merci_FR/s00076767n

Part 6: Conclusion

Acknowledgements

- ERC
- Maud Ehrmann, David Jurgens, Andrea Moro and Mohammad Taher Pilehvar for help with slides

Thanks or...





SAPIENZA
UNIVERSITÀ DI ROMA

Linguistic Computing Laboratory
<http://lcl.uniroma1.it>

